

Bowdoin College

## Bowdoin Digital Commons

---

Honors Projects

Student Scholarship and Creative Work

---

2023

### Una Comprensione Computazionale della Psiche Emotiva e Ordine nelle Ballate del *Decameron*: Stilometria e Elaborazione del Linguaggio Naturale

Nothando Khumalo  
*Bowdoin College*

Follow this and additional works at: <https://digitalcommons.bowdoin.edu/honorsprojects>



Part of the [European Languages and Societies Commons](#), [Italian Linguistics Commons](#), [Italian Literature Commons](#), and the [Medieval Studies Commons](#)

---

#### Recommended Citation

Khumalo, Nothando, "Una Comprensione Computazionale della Psiche Emotiva e Ordine nelle Ballate del *Decameron*: Stilometria e Elaborazione del Linguaggio Naturale" (2023). *Honors Projects*. 435.  
<https://digitalcommons.bowdoin.edu/honorsprojects/435>

This Open Access Thesis is brought to you for free and open access by the Student Scholarship and Creative Work at Bowdoin Digital Commons. It has been accepted for inclusion in Honors Projects by an authorized administrator of Bowdoin Digital Commons. For more information, please contact [mdoyle@bowdoin.edu](mailto:mdoyle@bowdoin.edu), [a.sauer@bowdoin.edu](mailto:a.sauer@bowdoin.edu).

Una Comprensione Computazionale della Psiche Emotiva e Ordine nelle Ballate del *Decameron*:  
Stilometria e Elaborazione del Linguaggio Naturale

An Honors Paper for the Department of Romance Languages and Literatures

By Nothando Khumalo

Bowdoin College, 2023

©2023 Nothando Khumalo

## Indice Sommari

|   |           |
|---|-----------|
| <b>RINGRAZIAMENTI</b> .....   | <b>3</b>  |
| <b>UNA NOTA SULLE CITAZIONI DEL <i>DECAMERON</i></b> .....  | <b>4</b>  |
| <b>PREFAZIONE</b> .....   | <b>5</b>  |
| <b>INTRODUZIONE</b> .....   | <b>6</b>  |
| <i>INTRODUZIONE AL DECAMERON</i> .....  | 6         |
| <i>IL DECAMERON E LE DIGITAL HUMANITIES</i> .....   | 10        |
| <i>UNO STUDIO PRECEDENTE</i> .....  | 12        |
| <i>INTRODUZIONE ALLA STILOMETRIA</i> .....  | 15        |
| <i>METODOLOGIA DELLA STILOMETRIA</i> .....  | 18        |
| <i>LA PSICHE EMOTIVA E LE BALLATE</i> .....   | 19        |
| <i>INTRODUZIONE ALL'ELABORAZIONE DEL LINGUAGGIO NATURALE</i> .....  | 22        |
| <i>CLASSIFICAZIONE DEL TESTO</i> .....  | 23        |
| <i>BAG OF WORDS</i> .....   | 24        |
| <i>WORD2VEC</i> .....   | 25        |
| <i>LA REGRESSIONE LINEARE</i> .....   | 26        |
| <i>IL TEOREMA DI BAYES</i> .....  | 28        |
| <i>ORDINE E LE BALLATE</i> .....  | 29        |
| <b>STILOMETRIA (PSICHE EMOTIVA)</b> .....   | <b>32</b> |
| <i>DENDROGRAMMA: VISUALIZZAZIONI DELLA STILOMETRIA</i> .....  | 32        |
| <i>LA BALLATA DI PANFILO</i> .....  | 39        |
| <i>LA BALLATA DI FILOSTRATO</i> .....   | 43        |
| <i>PANFILO E FILOSTRATO: PSICHE EMOTIVA TRASCENDE IL GENERE</i> .....   | 48        |
| <i>RAGGRUPParsi CON FIAMMETTA E FILOMENA FA CHIAREZZA SULLE EMOZIONI DI PANFILO</i> .....   | 50        |
| <i>GRUPPO 2: LA LINEA TRA FELICITÀ E TRISTEZZA</i> .....  | 52        |
| <i>RIASSUNTO</i> .....  | 57        |
| <b>ELABORAZIONE DEL LINGUAGGIO NATURALE (ORDINE)</b> .....  | <b>58</b> |
| <i>METODOLOGIA DI CLASSIFICAZIONE DEL TESTO</i> .....   | 58        |
| <i>USANDO IL ELN PER MOSTRARE L'ORDINE NEL DECAMERON</i> .....  | 60        |
| <i>SCELTA DI RIMUOVERE LA LEMMATIZZAZIONE E IL TAGGING DI PARTE DEL DISCORSO DAL PROCESSO DI PRE-ELABORAZIONE DEL TESTO</i> ..... | 62        |
| <i>SCELTA DI WORD2VEC E REGRESSIONE LINEARE</i> .....   | 64        |
| <i>PREDIZIONE DI AUTORE DELLE BALLATE</i> .....   | 69        |
| <i>ANALISI DI PANFILO</i> .....   | 70        |
| <i>ANALISI DI NEIFILE</i> .....   | 71        |
| <i>ANALISI DI FILOSTRATO</i> .....  | 73        |
| <i>ANALISI DI LAURETTA</i> .....  | 75        |
| <i>RIASSUNTO</i> .....  | 76        |
| <b>CONCLUSIONE</b> .....  | <b>78</b> |
| <i>RIASSUNTO DEI RISULTATI</i> .....  | 78        |
| <i>IMPATTO DELLO STUDIO SUL DECAMERON</i> .....   | 79        |
| <i>MIGLIORAMENTO</i> .....  | 80        |
| <i>IL MIO CAMMINO VERSO IL DECAMERON</i> .....  | 81        |
| <b>APPENDICE</b> .....  | <b>85</b> |
| <b>RIFERIMENTI</b> .....  | <b>86</b> |

## **Ringraziamenti**

Grazie mille al dipartimento di italiano per avermi insegnato l'italiano. Senza di voi non riuscirei a leggere questo meraviglioso libro in italiano. Anche questo progetto non sarebbe possibile senza la guida e il supporto del mio consulente onorario, la professoressa Crystal Hall. Grazie per aver camminato con me e avermi insegnato tutto quello che sai sulla letteratura italiana medievale e le discipline umanistiche digitali. Un enorme grazie ai professori italiani e dell'informatica che mi hanno insegnato la lingua che mi ha avvicinato a me stesso: la professoressa Anna Rein, la professoressa Davida Gavioli, la professoressa Allison Cooper, e il professore David Byrd. Un ringraziamento speciale al lettore della mia tesi, la professoressa Anna Rein.

## **Una Nota sulle Citazioni del *Decameron***

Nel citare il testo italiano del *Decameron* mi sono affidato a Giovanni Boccaccio, *Decameron*, ed. Vittore Branca (Milano, Mondari, 1985). I passaggi del *Decameron* non sono identificati dal numero di pagina acquistato da Day, novella e sezione.

## Prefazione

“These and earlier proofs are enough to show that poets, though poor, have been wise, illustrious, and their fame lives on for eternity, while the wealth and names of jurists vanish like smoke into thin air. By the same course of reasoning, I have shown also that this work of mine, if poetry is worth anything, is least of all superfluous, nor were my labors in compiling it vain.”<sup>1</sup>

---

<sup>1</sup> Osood, *Boccaccio on Poetry: Being the Preface and the Fourteenth and Fifteenth Books of Boccaccio's Genealogia Deorum Gentilium in an English Version with Introductory Essay and Commentary*, 82:28.

## Introduzione

### *Introduzione al Decameron*

Il *Decameron* è una delle opere più celebri della letteratura italiana e una pietra miliare del canone letterario occidentale. Scritto dall'autore italiano Giovanni Boccaccio nel XIV secolo, è una raccolta di 100 racconti che sono interconnessi da una cornice narrativa.<sup>2</sup> È ampiamente considerato un capolavoro di narrazione, umorismo e arguzia e ha avuto un profondo impatto sulla letteratura e sulla cultura nel corso dei secoli.

Boccaccio scrisse il *Decameron* sulla scia della peste nera, una pandemia che devastò l'Europa nel XIV secolo, uccidendo milioni di persone.<sup>3</sup> Il libro è ambientato a Firenze, in Italia, durante lo scoppio della peste e segue le avventure di un gruppo di dieci giovani che fuggono dalla città per sfuggire alla malattia. Trascorrono dieci giornate in una villa di campagna, intrattenendosi a vicenda con storie d'amore, tragedie, arguzie e inganni. Il titolo del libro deriva dalle parole greche "deka" (dieci) e "hemera" (giornata), che si riferiscono alle dieci giornate che i personaggi trascorrono nella villa.

Il *Decameron* di Boccaccio fu un significativo allontanamento dalle tradizioni letterarie del suo tempo. L'era medievale era dominata dai testi religiosi e la narrazione era tipicamente riservata a sermoni e agiografie. Il *Decameron* era rivoluzionario nella sua attenzione ai temi secolari e umanistici e nel suo uso dell'italiano volgare invece del latino, la lingua della Chiesa e dell'aristocrazia. È stato uno dei primi esempi di letteratura italiana e ha svolto un ruolo cruciale nello sviluppo della lingua italiana.

---

<sup>2</sup> Cervigni, "The Decameron's Ballads and Emilia's Happy Song," 133.

<sup>3</sup> Cantor, *In the Wake of the Plague*, 4.

Boccaccio nacque nel 1313 a Firenze, in Italia, da una famiglia di ricchi mercanti.<sup>4</sup> Ha ricevuto un'istruzione classica e ha studiato legge, ma la sua vera passione era la letteratura. Ha scritto poesie e prosa e le sue opere hanno abbracciato una vasta gamma di generi, tra cui romanticismo, allegoria, biografia e narrativa storica. Fu amico e contemporaneo di molti dei grandi letterati del suo tempo, tra cui Petrarca e Dante Alighieri.

Boccaccio visse un periodo tumultuoso della storia italiana, segnato da sconvolgimenti politici e sociali, guerre e peste nera. Il XIV secolo vide il declino del sistema feudale e l'ascesa di città-stato, come Firenze, che divennero centri di commercio, cultura e ricerca intellettuale.<sup>5</sup> Il Rinascimento cominciava a prendere forma e gli ideali umanistici dell'antichità classica venivano riscoperti e celebrati. Il *Decameron* riflette questo mutevole panorama culturale, con la sua enfasi sull'individualismo, il secolarismo e il valore della vita umana.

Il *Decameron* ebbe un successo immediato fin dalla sua pubblicazione nel 1353.<sup>6</sup> Fu ampiamente letto e copiato in tutta Europa e influenzò innumerevoli scrittori e artisti nei secoli successivi. È stato tradotto in dozzine di lingue e la sua influenza può essere vista in opere diverse come *Canterbury Tales* di Chaucer, opere di Shakespeare e film di Federico Fellini.<sup>7</sup>

Il *Decameron* è diviso in dieci giornate, e ogni giornata contiene dieci storie, raccontate da un diverso membro del gruppo. Le storie coprono una vasta gamma di temi e generi, tra cui storie d'amore, tragedia, commedia e satira. Riflettono la diversità dell'esperienza umana e le complessità della psiche umana.

---

<sup>4</sup> Falvo, "Ritual and Ceremony in Boccaccio's 'Decameron,'" 144.

<sup>5</sup> Cantor, *In the Wake of the Plague*, 8.

<sup>6</sup> Cervigni, "The Decameron's Ballads and Emilia's Happy Song."

<sup>7</sup> Cottino-Jones, "The Mode and Structure of Tragedy in Boccaccio's Decameron (IV, 9)."



Il *Decameron* è un capolavoro della letteratura italiana e una testimonianza del potere duraturo della narrazione. È un'opera che celebra la diversità dell'esperienza umana e la complessità della psiche umana. Riflette il mutevole panorama culturale del XIV secolo e l'ascesa degli ideali umanistici che avrebbero plasmato la civiltà occidentale nei secoli a venire. Il *Decameron* di Boccaccio è un'opera che ha ispirato e influenzato innumerevoli scrittori e artisti, e continua ad affascinare i lettori fino ad oggi.

La struttura di base di una giornata nel *Decameron* è la seguente:

1. Introduzione: ogni giornata inizia con una breve descrizione dell'ambiente e del tempo.
2. Tema: ogni giornata ha un tema o un argomento specifico, scelto dalla regina o dal re della giornata. Il tema può spaziare dall'amore e dalla cavalleria all'inganno e al tradimento.
3. Storytelling: i membri del gruppo raccontano a turno storie legate al tema del giornata. Ogni persona racconta una storia al giornata, per un totale di 10 storie al giornata.
4. Intermezzi: tra una storia e l'altra, il gruppo si impegna in varie attività come musica, balli e giochi.
5. Conclusione: Alla fine della giornata, la regina o il re dichiara il vincitore del concorso di narrazione e un nuovo sovrano viene selezionato dal precedente re o regina. Il nuovo sovrano sceglie qualcuno per cantare una ballata e il gruppo si ritira per la notte.

Nel complesso, la struttura di una giornata del *Decameron* è progettata per fornire un quadro per la narrazione, consentendo anche pause e socializzazione tra i membri del gruppo.

| <b>MEMBRO DELLA<br/>BRIGATA</b> | <b>ETÀ</b>   | <b>TEMA DEL GIORNATA</b>  |
|---------------------------------|--------------|---|
| <b>Pampinea</b>                 | <b>18-21</b> | <b>• (1)</b>  |
| <b>Fiammetta</b>                | <b>18-21</b> | <b>“gli amanti attraversano disastri prima che il loro amore finisca in buona fortuna”<sup>8</sup> (5)</b>                        |
| <b>Filomena</b>                 | <b>30-40</b> | <b>“chi, da diverse cose infestato, sia oltre alla speranza riuscito a lieto fine”<sup>9</sup> (2)</b>                            |
| <b>Emilia</b>                   | <b>18-21</b> | <b>• (9)</b>  |
| <b>Lauretta</b>                 | <b>18-21</b> | <b>“delle beffe che gli uomini fanno alle lor mogli”<sup>10</sup> (8)</b>   |
| <b>Neifile</b>                  | <b>30-40</b> | <b>“di chi alcuna cosa molto desiderata con industria acquistasse o la perdita recuperasse”<sup>11</sup> (3)</b>                  |
| <b>Elissa</b>                   | <b>18-21</b> | <b>“con pronta risposta o avvedimento fuggì perdita, pericolo o scorno”<sup>12</sup> (6)</b>                                      |
| <b>Filostrato</b>               | <b>18-21</b> | <b>“amori ebbero infelice fine”<sup>13</sup> (4)</b>  |
| <b>Panfilo</b>                  | <b>18-21</b> | <b>“di chi liberalmente ovvero magnificamente alcuna cosa operasse intorno a' fatti d'amore o d'altra cosa”<sup>14</sup> (10)</b> |
| <b>Dioneo</b>                   | <b>20-30</b> | <b>“beffe le quali o per amore o per salvamento di loro le donne hanno già fatte a' lor mariti”<sup>15</sup> (7)</b>              |

Tabella 1. I Membri della Brigata, le loro età e il tema del giornata che governano. Il numero della giornata in cui sono il re or la regina è indicato tra parentesi.

1. Pampinea - Viene descritta come una giovane donna ricca e bella, che è il capo della brigata.

<sup>8</sup> Boccaccio, *Decameron*, (4, Conclusione, 5)

<sup>9</sup> Boccaccio, *Decameron*, (1, Conclusione, 11)

<sup>10</sup> Boccaccio, *Decameron*, (7, Conclusione, 8)

<sup>11</sup> Boccaccio, *Decameron*, (2, Conclusione, 8)

<sup>12</sup> Boccaccio, *Decameron*, (5, Conclusione, 3)

<sup>13</sup> Boccaccio, *Decameron*, (3, Conclusione, 6)

<sup>14</sup> Boccaccio, *Decameron*, (9, Conclusione, 4)

<sup>15</sup> Boccaccio, *Decameron*, (6, Conclusione, 6)

2. Fiammetta - Viene descritta come una giovane donna bella e affascinante, ben istruita in letteratura e musica.
3. Filomena - Viene descritta come una donna saggia e matura, abile nella narrazione.
4. Emilia - Viene descritta come una giovane donna modesta e virtuosa, colta e eloquente oppure che parla bene.
5. Lauretta - Viene descritta come una giovane donna bella e spiritosa, abile nel canto e nel suonare strumenti musicali.
6. Neifile - Viene descritta come una donna graziosa e dignitosa, esperta nelle relazioni amorose.
7. Elissa - Viene descritta come una giovane donna bella e intelligente, esperta di filosofia e teologia.
8. Filostrato - Viene descritto come un giovane bello e coraggioso, esperto nelle arti marziali e nell'equitazione.
9. Panfilo - Viene descritto come giovane cortese e grazioso, abile nella musica e nel ballo.
10. Dioneo - Viene descritto come un giovane spiritoso e irriverente, abile nella narrazione e nel canto.

### *Il Decameron e l'Umanistica Digitale*

Il *Decameron* è un testo interessante di umanistica digitale perché offre l'opportunità di analizzare un testo medievale utilizzando tecnologie e metodologie moderne. In quanto opera letteraria, fornisce una panoramica della società, della cultura e dei valori italiani medievali. In quanto argomento per umanistica digitale, il *Decameron* offre una ricca fonte di dati che possono essere analizzati utilizzando strumenti e tecniche computazionali.

Il *Decameron* porta avanti un insieme complesso di sfide e opportunità. Il testo è stato oggetto di ampi studi e interpretazioni nel corso dei secoli, ma gli approcci delle umanistiche digitale offrono nuovi modi per esplorare il testo e il suo contesto. Il *Decameron* presenta sfide per l'analisi perché contiene molteplici livelli di significato e interpretazione. La lettura del testo offre già al lettore un compito impegnativo. L'analisi computazionale è altrettanto impegnativa. Allo stesso tempo, il *Decameron* offre opportunità per l'analisi computazionale perché, in quanto testo di grandi dimensioni, contiene una grande quantità di dati che possono essere analizzati utilizzando strumenti e tecniche dell'umanistica digitale. Applicando metodi computazionali al testo, gli studiosi di discipline umanistiche digitali possono esplorare questioni di lingua, stile, paternità e contesto culturale, nonché scoprire modelli e relazioni precedentemente non riconosciuti all'interno del testo.

Le ballate del *Decameron* sono argomenti interessanti per la ricerca nell'umanistica digitale per diversi motivi. In primo luogo, sono parte integrante della struttura e della narrazione del *Decameron*, in quanto forniscono un quadro per le storie di ogni giornata e aiutano a stabilire i temi generali dell'opera. In secondo luogo, le ballate stesse sono ricche di significato letterario e culturale, attingendo a un'ampia gamma di tradizioni musicali e poetiche di tutta l'Europa medievale. Infine, le ballate offrono una finestra unica sulle vite e le esperienze emotive della brigata, permettendoci di esplorare i modi in cui hanno espresso i loro sentimenti ed esperienze attraverso la musica e la poesia. Utilizzando strumenti digitali come la stilometria e l'elaborazione del linguaggio naturale, possiamo acquisire nuove conoscenze sulla struttura e sul significato delle ballate, nonché sul loro posto nel contesto più ampio della letteratura e della

cultura medievali. In questo studio vengono proposti i due scopi principali delle ballate nel *Decameron*. In primo luogo, fungono da finestra sulla psiche emotiva di ogni cantante. In secondo luogo, mirano a ristabilire l'ordine alla fine di ogni giornata. Utilizzando la stilometria e l'elaborazione del linguaggio naturale, è possibile determinare lo stato emotivo della brigata e l'ordine che le ballate portano al testo.

### *Uno Studio Precedente*

Nell'articolo “‘Tecnologica cosa’: Modeling Storyteller Personalities in Boccaccio’s *Decameron*” A. Feder Cooper, Marilyn Migiel e altri ricercatori della Cornell University esplorano come gli strumenti delle discipline umanistiche digitali possono essere utilizzati per studiare il *Decameron* di Boccaccio.<sup>16</sup> Nello specifico, gli autori si concentrano sulla questione se i diversi narratori nel testo esibiscano personalità distinte. Curano un set di dati basato sull'edizione autorevole del testo e utilizzano metodi di classificazione supervisionati per prevedere i narratori in base alle storie che raccontano. Dimostrano anche che la modellazione degli argomenti può estrarre i "profili" del narratore tematico. Mentre la ricerca su *The Canterbury Tales* si è impegnata sia con le storie che con i narratori, la ricerca letteraria sul *Decameron* ha avuto la tendenza ad affrontare temi e storie piuttosto che l'identità del narratore. Gli autori cercano di colmare questa lacuna utilizzando strumenti computazionali per ricavare modelli dal testo che potrebbero essere rimasti sfuggenti agli studiosi.

Questo studio evidenzia diverse sfide per la ricerca nelle discipline umanistiche digitali, in particolare per un testo scritto in italiano medievale con poche risorse di modellazione

---

<sup>16</sup> Cooper et al., “‘Tecnologica Cosa’,” 147.

linguistica. Nonostante queste sfide, questo approccio per fornire approfondimenti su come gli strumenti delle discipline umanistiche digitali possono essere utilizzati per studiare un testo come il *Decameron*. Hanno eseguito tre test principali: (1) una regressione logica per determinare se i membri della brigata avevano voci distinguibili, (2) modellazione degli argomenti per vedere se alcuni argomenti erano più prevalenti in certa novelle dei membri della brigata. Quest'ultimo test è stato ripetuto per vedere se qualche argomento fosse coerente tra i membri della brigata dello stesso sesso.

Gli autori dello studio hanno elaborato un modello di regressione logistica per classificare i narratori del testo nel *Decameron*. Gli autori hanno scoperto che solo due narratori, entrambi uomini, hanno costantemente ottenuto punteggi migliori rispetto a quelli casuali nel loro compito di classificazione: Panfilo e Dioneo. La migliore interpretazione di Dioneo non è stata sorprendente, poiché ha avuto il privilegio speciale di deviare dal tema narrativo della Giornata. Tuttavia, era meno chiaro il motivo per cui Panfilo si è distinto nei risultati.

In un successivo esperimento, gli autori hanno pre-elaborato il *Decameron* in modo che contenesse solo le 100 parole usate più di frequente nel vocabolario. Questo modello ha potuto identificare Fiammetta, una delle sette donne, meglio che a caso, oltre a Panfilo e Dioneo. Gli autori hanno anche verificato i risultati della loro classificazione rieseguendo il test solo con le parole più rilevanti e importanti in un corpus. Hanno scoperto che, nonostante le basse prestazioni del classificatore, le differenze lessicali tra i narratori erano interpretabili. Ad esempio, Neifile ha ottenuto un punteggio alto per le parole che significano onorabilità, mentre Filostrato ha ottenuto un punteggio basso per le parole riguardanti l'amore e la speranza. Nel

complesso, l'esperimento degli autori ha mostrato che ci sono differenze lessicali identificabili tra i narratori nel *Decameron*, sebbene il compito di classificazione fosse impegnativo. Gli autori suggeriscono che i loro risultati potrebbero portare a ulteriori indagini sul motivo per cui alcuni narratori erano più identificabili di altri.

Oltre alla classificazione dei membri della brigata, gli autori descrivono il loro uso dell'allocatione di Dirichlet latente (ADL) per modellare ogni novella come una distribuzione di argomenti e raggruppare i risultati per narratore per vedere se le distribuzioni degli argomenti della novella sono distinguibili per ciascuno dei 10 membri della brigata. Gli autori vedono le distribuzioni degli argomenti per narratore come "profili" del narratore che possono indicare caratteristiche tematiche uniche di particolari membri della brigata. Gli autori hanno osservato 20 argomenti: agrario, morte/dolore, famiglia/matrimonio/figli, fortuna, gentilezza/nobiltà, amore, mercantile, religione/frati, religione/frequentare, religione/abramo, religione/monache, marittimo e urbano. Quindi convalidano gli argomenti rimanenti a livello di novella, tracciando una mappa termica in cui ogni riga è una distribuzione degli argomenti toccati dell'ogni membri della brigata. Questa mappa termica consente loro di controllare a campione se un particolare novella ha una ragionevole distribuzione degli argomenti. Osservano celle nella mappa termica che mostrano una presenza unica e molto ponderata per un argomento che è relativamente assente per ciascuno degli altri nove narratori, come religione/peccato per Pampinea, mercantile per Dioneo e marinara per Panfilo.

Gli autori eseguono anche un'analisi simile, raggruppando le distribuzioni degli argomenti in modo più grossolano per genere del narratore invece che per singolo narratore. Osservano che gli

uomini discutono di temi mercantili molto più delle donne e le donne discutono dell'amore più degli uomini. Sugeriscono che forse i risultati mercantili non sono sorprendenti data la capacità illimitata degli uomini di partecipare agli sforzi economici, ma il risultato riguardante l'amore è in qualche modo sorprendente dato lo scopo dichiarato dell'Autore nel Proemio di alleviare la sofferenza delle donne innamorate, e si dice che i tre uomini siano innamorati di tre delle donne della brigata.

Le loro scoperte mi sono state utili per decidere a quali test computazionali sottoporre il testo. In particolare, la loro determinazione della voce di Panfilo come argomento interessante si è rispecchiata nelle mie stesse scoperte. Usando questo studio come guida e dopo aver testato vari metodi computazionali, ho deciso che la stilometria e l'elaborazione della lingua naturale sarebbero state molto utili per fornire maggiori informazioni sul valore retorico e strutturale delle ballate del *Decameron*.

### *Introduzione alla Stilometria*

Pensa a come parla ciascuno dei tuoi amici. Riusciresti a identificare tutti i tuoi amici anche se non li vedevi, tutte le loro voci suonano allo stesso modo e parlano tutti la stessa lingua? Sebbene tutti i tuoi amici abbiano lo stesso lessico a loro disposizione, non è irragionevole suggerire che saresti in grado di identificarli con successo senza troppi errori. Questo perché il tono, la scelta delle parole e altri modelli di discorso sono diversi per tutti.<sup>17</sup> I modelli di linguaggio di una persona sono spesso rappresentativi della sua età, sesso, sessualità, religione e razza. A un livello più profondo, la parola è rappresentativa della personalità di una persona, della sua psiche

---

<sup>17</sup> Jensen and DiTiberio, "Personality and Individual Writing Processes," 287.



emotiva interiore. Gli schemi linguistici unici di ogni persona rappresentano il suo processo di pensiero interiore. Conclusioni simili si potrebbero trarre riguardo al testo scritto perché nella sua forma più pura è scritto pensando. Alcuni ricercatori arrivano persino a caratterizzare la scrittura come parte integrante dello sviluppo di una personalità sfumata.<sup>18</sup> Il cervello umano può eseguire questo complesso processo di identificazione con testi e parole naturali senza troppi sforzi. Può facilmente identificare uno stiloma o “a linguistic fingerprint that can be measured, is largely unconscious, and is constant.”<sup>19</sup> Un computer può fare lo stesso?

La stilometria computazionale tenta di fare esattamente questo. La stilometria, o metrica dello stile, si riferisce a un'analisi quantitativa e statistica dello stile letterario o di scrittura.<sup>20</sup> Viene spesso utilizzata dai ricercatori per il "riconoscimento dell'autore", "l'identificazione dell'autore" e "l'attribuzione dell'autore".<sup>21</sup> Ma come è possibile? Al di fuori dell'argomento, un computer può davvero discernere correttamente la differenza tra diversi parlanti? Di quali classificazioni dispone un computer per condurre un'analisi statistica significativa? Non c'è molto spazio per la creatività quando si tratta di grammatica, quindi cosa c'è di così particolare nei singoli parlanti di una lingua? Quali conoscenze utili si possono estrarre da un testo? Daelamans ha identificato tre diversi tipi di conoscenza che possono essere ottenuti da un testo:

The three types of knowledge we distinguish that can be extracted from text are: (i) objective knowledge (answering the who, what, where, when, ... questions), (ii) subjective

---

<sup>18</sup> Jensen and DiTiberio, 286.

<sup>19</sup> van Halteren et al., “New Machine Learning Methods Demonstrate the Existence of a Human Stylome,” 3.

<sup>20</sup> van Halteren et al., 6.

<sup>21</sup> van Halteren et al., 3.

knowledge (who has which opinion about what?), and (iii) metaknowledge (what can we extract about the text apart from its contents, mainly about its author?).<sup>22</sup>

La conoscenza più utile per uno studio della psiche emotiva interiore di una persona è la terza. La metaconoscenza include informazioni direttamente collegate alla psicologia e alla sociologia dell'autore o del parlante. Alcuni fattori che i ricercatori hanno trovato essere particolarmente importanti sono la lunghezza di una frase, la posizione delle parole in una frase, parte del discorso e la frequenza delle parole accanto a un'altra parola specifica.<sup>23</sup> La lunghezza di una frase in cui è possibile trovare una parola viene memorizzata come valore intero. La posizione di una parola in una frase si riferisce alla posizione rispetto alle altre parole. Consideriamo una frase semplice: "I speak Italian." La parola "speak" si trova accanto alla parola "I" e "Italian". A questa relazione è associato anche un valore di frequenza. Se questa frequenza si è verificata più di una volta, il conteggio della frequenza aumenta di conseguenza.

Ho scelto di rimuovere le stop word da tutti i test stilometrici che ho condotto. Tuttavia, Matthew Jockers, un pioniere nel campo delle discipline umanistiche digitali, di solito include stop word nei suoi test:

Many years of authorship attribution research have taught us that the most effective way to distinguish between the text of one author and another is by comparing the different usages of high frequency features in their writing. High frequency features include words such as *the*, *of*, *and*, and *to*, as well as, in some studies, marks of punctuation and even common bigrams, such as *of the*.<sup>24</sup>

---

<sup>22</sup> van Halteren et al., 2.

<sup>23</sup> van Halteren et al., 10.

<sup>24</sup> Jockers, *Text Analysis with R for Students of Literature*, 109.

Tuttavia, la sua scelta di includere stop word ad alta frequenza come indicatore di differenza stilistica era in diretta contraddizione con la mia scelta di ignorarle. Ho deciso di continuare ad ignorarle perché la frequenza degli articoli in italiano è molto più alta che in inglese. La frequenza degli articoli in italiano non rappresenta una scelta stilistica, ma riflette semplicemente un principio chiave della lingua.

### *Metodologia della stilometria*

La distanza euclidea è una misura della distanza tra due punti in uno spazio multidimensionale. Nell'analisi stilometrica, questa distanza viene utilizzata per misurare la somiglianza o la dissomiglianza tra due testi. Questa tecnica si basa sull'idea che ogni testo può essere rappresentato come un vettore in uno spazio ad alta dimensione, dove ogni dimensione corrisponde a una caratteristica linguistica specifica, come la frequenza delle parole, la lunghezza delle parole o la lunghezza della frase. I test stilometrici possono utilizzare varie misure di distanza per calcolare la somiglianza o la dissomiglianza tra due testi, inclusa la distanza euclidea. In questo caso, la distanza tra due testi viene calcolata come radice quadrata della somma delle differenze al quadrato tra i loro valori di caratteristica in ciascuna dimensione. In altre parole, la distanza euclidea misura la lunghezza della retta che collega i due vettori nello spazio ad alta dimensione.

Ad esempio, supponiamo di avere due testi A e B, ciascuno rappresentato come un vettore con due dimensioni: frequenza delle parole e lunghezza della frase. La rappresentazione vettoriale del testo A potrebbe essere [0.2, 12], indicando che il 20% delle sue parole sono una parola particolare e la sua lunghezza media della frase è di 12 parole. La rappresentazione vettoriale del

testo B potrebbe essere [0.3, 8], indicando che il 30% delle sue parole sono la stessa parola particolare e la sua lunghezza media della frase è di 8 parole. Per calcolare la distanza euclidea tra A e B, calcoliamo prima le differenze al quadrato tra i loro valori di caratteristica in ciascuna dimensione. La distanza euclidea risultante tra A e B è 4.001.

I test stilometrici possono utilizzare questa misura di distanza per distinguere i testi in base alle loro caratteristiche stilistiche. Ad esempio, un test stilometrico potrebbe analizzare la frequenza delle parole, la lunghezza delle frasi o altre caratteristiche di due testi per determinare se sono stati scritti dallo stesso autore. Se la distanza euclidea tra i testi è piccola, suggerisce che sono stilisticamente simili e potrebbero essere stati scritti dallo stesso autore. Se la distanza euclidea è grande, suggerisce che i testi sono dissimili e probabilmente sono stati scritti da autori diversi.

### *La Psiche Emotiva e le Ballate*

Le ballate del *Decameron* sono un argomento interessante per l'analisi stilometrica nel testo. Le ballate sono identificate come poesie da ballare da mettere in musica. In quanto opere di poesia, svolgono il ruolo unico di esporre i desideri dei membri della brigata, desideri che non sono prontamente espressi nelle sezioni narrative del testo. Anche Giovanni Boccaccio ha scritto molto di poesia come “a science worthy of veneration” nel rivelare le emozioni umane.<sup>25</sup> Ho scelto di utilizzare una traduzione della *Genealogia Deorum Gentilium* di Boccaccio per rendere più accessibile il suo pensiero sulla poesia. L'unica versione esistente è stata tradotta da Charles G. Osgood della Princeton University nel 1929 e ripubblicata nel 1956.<sup>26</sup> Boccaccio parla sul ruolo della poesia:

---

<sup>25</sup> Osood, *Boccaccio on Poetry: Being the Preface and the Fourteenth and Fifteenth Books of Boccaccio's Genealogia Deorum Gentilium in an English Version with Introductory Essay and Commentary.*, 39

<sup>26</sup> Osood.:v.

This fervor of poesy is sublime in its effects: it impels the soul to a longing for the utterance; it brings forth strange and unheard-of creations of the mind; it arranges these meditations in a fixed order, adorns the whole composition with unusual interweaving of words and thoughts; and thus veils truth in a fair and fitting garment of fiction.<sup>27</sup>

Il termine “sublime” dimostra che Boccaccio credeva che la poesia lavori per aumentare la grandezza del pensiero umano. La frase "spinge l'anima a un desiderio" rivela che la poesia trasmette i desideri che sono parte integrante del proprio essere, desideri che sono spesso nascosti sotto la superficie. Inoltre, il termine “impel” suggerisce che il desiderio o le emozioni non sono scelti da una persona ma invece suggeriti dalla poesia. La poesia spinge le persone a trovare nuove profondità delle emozioni. L’espressione “unheard of creations of the mind” sottolinea questo sentimento e mostra la natura fluida e inventiva della mente umana. Boccaccio molto probabilmente credeva che la poesia fornisse agli esseri umani un mezzo per applicare un “fixed order” alle menti inquiete. L'uso della frase “interweaving of words and thoughts” allude all’idea che la poesia sia pensata in forma di parola. Tutta questa analisi mostra che Boccaccio ha posto la poesia su un alto piedistallo. Pertanto, si può suggerire che la poesia nel suo libro, le ballate, esprima anche la psiche emotiva interiore di ciascuno dei membri della brigata.

È importante ricordare lo sfondo del *Decameron*: la Peste Nera. La brigata non fuggì da Firenze solo per divertirsi ma anche per allontanarsi dalla morte che circondava la loro vita in città. Nel 1348, la peste bubbonica colpì Italia, Francia, e Medio Oriente.<sup>28</sup> A Firenze “infra 'l marzo e il prossimo luglio vegnente, ...oltre a centomilia creature umane si crede per certo dentro alle mura

---

<sup>27</sup> Osood, 82:39.

<sup>28</sup> Cantor, *In the Wake of the Plague*, 12.

della città” (I, 047).<sup>29</sup> In che modo questo evento traumatico può averli colpiti? Boccaccio offre uno spaccato del grado di tragedia che la peste portò a Firenze.

Secondo Boccaccio, la peste rappresentava l'inevitabile ira di Dio che si traduceva in un isolamento che spingeva i fortunati che erano sani a due linee d'azione principali.

E in quella non valendo alcuno senno né umano provvedimento, per lo quale fu da molte immondizie purgata la città da ufficiali sopra ciò ordinati e vietato l'entrarvi dentro a ciascuno infermo e molti consigli dati a conservazion della sanità, né ancora umili supplicazioni non una volta ma molte e in processioni ordinate, in altre guise a Dio fatte dalle divote persone, quasi nel principio della primavera dell'anno predetto orribilmente cominciò i suoi dolorosi effetti, e in miracolosa maniera, a dimostrare. (I, 009)<sup>30</sup>

E più avanti ancora ebbe di male: ché non solamente il parlare e l'usare cogli infermi dava a' sani infermità o cagione di comune morte, ma ancora il toccare i panni o qualunque altra cosa da quegli infermi stata tocca o adoperata pareva seco quella cotale infermità nel toccator trasportare. (I, 015)<sup>31</sup>

Dalle quali cose e da assai altre a queste simiglianti o maggiori nacquero diverse paure e immaginazioni in quegli che rimanevano vivi, e tutti quasi a un fine tiravano assai crudele, ciò era di schifare e di fuggire gl'infermi e le lor cose; e così facendo, si credeva ciascuno a se medesimo salute acquistare. (I, 019)<sup>32</sup>

Le persone sane sono fuggite dalla città per sfuggire alla malattia o sono rimaste per dedicarsi all'edonismo prima delle inevitabili morti. La brigata prende una via di mezzo tra queste due strade. Se la peste ha spinto il gruppo fuori dalla città che una volta chiamavano casa, deve aver

---

<sup>29</sup> Boccaccio and Branca, *Decameron*.

<sup>30</sup> Boccaccio and Branca.

<sup>31</sup> Boccaccio and Branca.

<sup>32</sup> Boccaccio and Branca.

spinto le loro menti in luoghi in cui non erano mai stati emotivamente. Un modo per esprimere questa emozione è attraverso la poesia o le ballate, che Boccaccio definisce come un tipo di poesia cantata in musica per ballare.

### *L'Introduzione all'Elaborazione del Linguaggio Naturale*

Ora che è stato stabilito che le ballate rivelano la voce unica di ciascuno dei personaggi, verrà utilizzata l'elaborazione del linguaggio naturale per verificare se le voci uniche dei membri della brigata sono mantenute o meno all'interno di ciascuna delle loro dieci storie. L'elaborazione del linguaggio naturale (ELN) esiste all'interno di una parte dell'informatica chiamata intelligenza artificiale. Più specificamente, il linguaggio naturale tenta di dare ai computer la capacità di comprendere testi scritti e parole pronunciate proprio come gli umani. Questo potrebbe sembrare simile alla stilometria perché lo è. Tuttavia, c'è una distinzione importante. La stilometria chiede al computer di raccogliere solo schemi, quindi può essere interpretata come una sorta di "apprendimento stupido". D'altra parte, l'elaborazione del linguaggio naturale è "apprendimento intelligente" perché al computer viene chiesto di interpretare il testo e di comprendere effettivamente ciò che viene scritto o detto. Buoni esempi sono i rilevatori di spam forniti dalla maggior parte dei servizi di posta elettronica. Questi programmi possono classificare le e-mail come spam proprio come può fare un essere umano. Usare il ELN per categorizzare il testo si chiama classificazione del testo.

### *La Classificazione del Testo*

Nell'apprendimento automatico, la classificazione si riferisce al tipo di problema in cui i dati devono essere classificati in una o più classi.<sup>33</sup> I dati possono essere testo, un'immagine, parole pronunciate o numeri. La classificazione del testo è il caso che si occupa esplicitamente di testi e deve collocare uno o più testi di input in un insieme di classi predeterminate. Questi testi possono essere un singolo carattere come “a.”, una stringa di caratteri, o una parola, come “ciao”, una frase, o un intero libro. La lunghezza è arbitraria ed è determinata dall'utente. I problemi di classificazione del testo possono essere ulteriormente classificati in base al numero di classi possibili. La classificazione binaria implica una classificazione che può collocare un testo in solo due possibili categorie. L'identificazione multiclasse può collocare un testo in tre o più classi possibili. Per la classificazione binaria e multiclasse, un testo deve essere inserito in una sola categoria. Ad esempio, se stessimo classificando un'e-mail come non spam o spam, dovrebbe essere classificata come spam o non spam. Non c'è via di mezzo.

Si noti che è possibile che un dato testo sia classificato in più di una classe. Questo caso è possibile solo se un problema di classificazione del testo è multietichetta. La classificazione multietichetta consente di classificare un testo in più di una classe. Un testo può non essere inserito in nessuna classe, in una classe o in più di una classe. Questo è distinto dalla classificazione binaria e multiclasse. Per attribuire un autore a ciascuna delle ballate, utilizzeremo la classificazione multiclasse. L'attribuzione dell'autore è un uso comune della classificazione del testo ed è molto comune nel campo degli studi letterari. Per completare la classificazione, il testo naturale del *Decameron* deve essere trasformato in dati numerici che il

---

<sup>33</sup> Vajjala, “4. Text Classification - Practical Natural Language Processing [Book].”



computer possa comprendere. Esistono due classi principali di metodi: metodi basati sulla frequenza, come Bag-of-Words e metodi basati sull'incorporamento, come word2vec.

### *Bag-of-Words*

Il modello Bag-of-Words (BoW) è una tecnica utilizzata nel campo dell'Elaborazione della lingua naturale (ELN) per rappresentare un documento come un insieme di parole senza considerare l'ordine in cui compaiono. In pratica, il modello BoW funziona nel seguente modo:

1. Si prende un insieme di documenti e si crea un vocabolario unico, che contiene tutte le parole che compaiono nei documenti.
2. Si rappresenta ogni documento come un vettore che contiene il conteggio delle parole presenti nel documento, utilizzando il vocabolario creato nel passo precedente.
3. Si ignora completamente l'ordine in cui le parole compaiono nel documento e si considera solo il conteggio di ciascuna parola.

Ad esempio, supponiamo di avere due frasi: "Mario mangia la pizza" e "Luca mangia la pasta".

Possiamo creare un vocabolario contenente tutte le parole uniche presenti nelle frasi: ["Mario", "mangia", "la", "pizza", "Luca", "pasta"]. Quindi, rappresentiamo ogni frase come un vettore che contiene il conteggio di ciascuna parola presente nella frase, utilizzando il vocabolario creato:

"Mario mangia la pizza" diventa [1, 1, 1, 1, 0, 0]. "Luca mangia la pasta" diventa [0, 1, 1, 0, 1, 1].

Questo tipo di rappresentazione dei documenti è utile in molti contesti di ELN, come ad esempio la classificazione dei documenti. In particolare, il modello BoW permette di rappresentare un testo in modo semplice ed efficiente, senza dover considerare la struttura sintattica e semantica

del testo. Tuttavia, è importante sottolineare che il modello BoW ha alcune limitazioni, come la perdita di informazioni sulla posizione e sull'ordine delle parole all'interno dei documenti.

### *Word2vec*

Word2Vec è un modello di rappresentazione del linguaggio naturale utilizzato per catturare la semantica delle parole attraverso la costruzione di vettori di parole densi. Il modello Word2Vec utilizza una rete neurale artificiale per imparare la rappresentazione vettoriale di parole in un corpus di testi. In pratica, il modello Word2Vec analizza il contesto in cui una parola appare nel corpus e utilizza questa informazione per costruire un vettore di parole che cattura la relazione semantica tra le parole. Una volta addestrato, il modello Word2Vec può essere utilizzato per calcolare la somiglianza semantica tra le parole. In particolare, la similarità tra due parole può essere misurata come la distanza tra i loro vettori di parole. Le parole che sono semanticamente simili avranno vettori di parole più vicini tra loro nello spazio vettoriale.

Ad esempio, supponiamo di avere un corpus di testo contenente la frase "Mario mangia la pizza". Il modello Word2Vec può essere utilizzato per costruire un vettore di parole per ogni parola del corpus, come ad esempio  $[0.5, 0.2, 0.7, -0.3]$  per la parola "pizza". Questi vettori di parole possono poi essere utilizzati per calcolare la similarità semantica tra le parole.

Il modello Word2Vec è stato utilizzato in molti contesti di ELN, come ad esempio la classificazione dei documenti, la traduzione automatica e la generazione di testo. In particolare, il modello Word2Vec ha dimostrato di essere particolarmente efficace nel catturare la semantica delle parole in diverse lingue, compreso l'italiano. Una volta che i dati sono stati creati, devono essere appresi da un modello di apprendimento. Esistono due modelli principali: la regressione lineare e il classificatore ingenuo Bayes.

### *La Regressione Lineare*

La regressione lineare è un metodo statistico utilizzato per analizzare la relazione tra una variabile dipendente (o di output) e una o più variabili indipendenti (o di input). Nel contesto dell'Elaborazione di Lingua Naturale (ELN), la regressione lineare viene spesso utilizzata per modellizzare la relazione tra le caratteristiche linguistiche di un testo (ad esempio il numero di parole, la lunghezza delle frasi, le frequenze delle parole, ecc.) e una variabile di output, come ad esempio il sentimento del testo o la sua categoria. In pratica, una volta che si è raccolto un set di dati che includono sia le caratteristiche linguistiche di un testo che la sua variabile di output, si può utilizzare la regressione lineare per costruire un modello che descriva la relazione tra queste due variabili. Il modello utilizza le caratteristiche linguistiche come variabili indipendenti e la variabile di output come variabile dipendente. Il processo di regressione lineare consiste nel trovare la migliore linea retta che si adatti ai dati, in modo che la differenza tra i valori previsti dal modello e i valori reali sia la più piccola possibile. Questo processo si basa sulla minimizzazione della funzione di errore, che rappresenta la differenza tra i valori previsti dal modello e i valori reali. Una volta che si è costruito il modello di regressione lineare, si può utilizzarlo per fare previsioni su nuovi testi. Ad esempio, si può usare il modello per prevedere il sentimento di un nuovo testo, basandosi sulle sue caratteristiche linguistiche.

Ecco una lista numerata dei passi da seguire per utilizzare la regressione lineare per la modellizzazione dei dati ELN:

1. Raccolta dei dati: si raccoglie un insieme di dati che includano le caratteristiche linguistiche di un testo (ad esempio, il numero di parole, la lunghezza delle frasi, le

frequenze delle parole, ecc.) e la sua variabile di output (ad esempio, il sentimento del testo o la sua categoria).

2. Preparazione dei dati: si organizzano i dati in modo che siano pronti per l'analisi.

Questo può includere la rimozione delle stop word (come "il", "la", "e", ecc.), la normalizzazione dei dati (come la conversione di tutte le parole in minuscolo), la codifica delle variabili categoriali in numeri, ecc.

3. Creazione di un modello di regressione lineare: utilizzando uno strumento di analisi statistica come R o Python, si crea un modello di regressione lineare utilizzando le caratteristiche linguistiche come variabili indipendenti e la variabile di output come variabile dipendente.

4. Valutazione del modello: si valuta la bontà del modello utilizzando diverse metriche, come il coefficiente di determinazione  $R^2$ , l'errore quadratico medio (EQM) o l'errore assoluto medio (EAM). Queste metriche aiuteranno a capire quanto bene il modello si adatta ai dati.

5. Utilizzo del modello per le previsioni: una volta che è stato creato un modello di regressione lineare accurato, si può utilizzarlo per fare previsioni su nuovi dati. Ad esempio, si può utilizzare il modello per prevedere il sentimento di un nuovo testo, basandosi sulle sue caratteristiche linguistiche.

### *Il Teorema di Bayes*

Il teorema di Bayes è un metodo statistico utilizzato per calcolare la probabilità di determinate condizioni data la presenza di prove fornite. Nel contesto del ELN, il teorema di Bayes viene spesso utilizzato per modellizzare la probabilità di appartenenza di un testo a una determinata classe, come ad esempio la classe dei documenti "spam" o dei documenti "non spam". Il modello di classificatore ingenuo Bayes è un modello di classificazione che utilizza il teorema di Bayes per calcolare la probabilità di appartenenza di un testo a una determinata classe, data la presenza di alcune caratteristiche linguistiche (come le parole o le frasi utilizzate nel testo). Il modello è chiamato "naive" (ingenuo) perché assume che le caratteristiche linguistiche siano indipendenti tra loro, anche se in realtà possono esserci delle dipendenze tra di esse. Il processo di classificatore ingenuo Bayes consiste nel calcolare la probabilità a posteriori di ogni classe (data l'insieme) di caratteristiche linguistiche presenti nel testo, utilizzando il teorema di Bayes. In pratica, ciò significa che il modello calcola la probabilità che il testo appartenga a ogni classe possibile e sceglie la classe con la probabilità maggiore come etichetta di classificazione del testo. Ecco una spiegazione dei passi coinvolti nell'utilizzo del classificatore ingenuo Bayes per il ELN:

1. Raccolta dei dati: si raccoglie un insieme di dati che includano testi etichettati con le rispettive classi (ad esempio, "spam" o "non spam").
2. Preparazione dei dati: si organizzano i dati in modo che siano pronti per l'analisi. Questo può includere la rimozione delle stop word (come "il", "la", "e", ecc.), la normalizzazione dei dati (come la conversione di tutte le parole in minuscolo), la codifica delle variabili categoriali in numeri, ecc.

3. Creazione di un modello di classificatore ingenuo Bayes: utilizzando uno strumento di analisi statistica come R o Python, si crea un modello di classificazione Bayesiana ingenua utilizzando le caratteristiche linguistiche come variabili indipendenti e le classi come variabile dipendente.
4. Valutazione del modello: valuta la bontà del modello utilizzando diverse metriche, come la precisione, il richiamo e l'F1-score. Queste metriche ti aiuteranno a capire quanto bene il modello si adatta ai dati.
5. Utilizzo del modello per la classificazione: una volta che è stato creato un modello di classificatore ingenuo Bayes accurato, si può utilizzarlo per classificare nuovi testi. Ad esempio, si può utilizzare il modello per classificare nuovi messaggi di posta elettronica come "spam" o "non spam".

I metodi computazionali delineati sopra possono essere usati per capire come le ballate aggiungono ordine al *Decameron*.

### *Ordine e Le Ballate*

Il *Decameron* ha una struttura narrativa fissa e la ELN testerà se questo senso dell'ordine continua o meno nello stile tra i membri della brigata nelle loro canzoni. Ognuno delle dieci giornate contiene una storia di ciascuno dei membri della brigata. L'argomento di ogni giornata è scelto dal re o dalla regina della giornata. Per trasferire il potere, ogni sovrano deve cantare una

canzone e poi scegliere il successore. Questa struttura narrativa crea un ordine molto chiaro nella storia. Cervigni concorda che questo è il ruolo delle canzoni. Pampinea, il membro più anziano della brigata, detta le regole del gruppo nella Prima Giornata:

... festevolmente viver si vuole, né altra cagione dalle tristizie ci ha fatte fuggire. / Ma per ciò che le cose che sono senza modo non possono lungamente durare, io, che cominciatrice fui de' ragionamenti da' quali questa così bella compagnia è stata fatta, pensando al continuar della nostra letizia, estimo che di necessità sia convenire esser tra noi alcuno principale, il quale noi e onoriamo e ubidiamo come maggiore, nel quale ogni pensiero stea di doverci a lietamente vivere disporre. (I, 94-95)<sup>34</sup>

La parola “festevolmente” stabilisce l'obiettivo generale del gruppo: la felicità! La brigata aveva lasciato Firenze per lasciarsi alle spalle la morte e la tristezza della città. La gioia che desiderano provare in campagna esiste in diretto contrasto con le “tristizie” che i membri avevano vissuto a Firenze. Dopo il trauma vissuto a Firenze, è comprensibile che la brigata sia attratta dall'ordine. La frase “le cose che sono senza modo non possono lungamente durare” sottolinea questo desiderio di ordine. “Le cose senza ordine” è un riferimento diretto alla morte insensata e alla sofferenza durante la peste. Pampinea si nomina protagonista principale della prima giornata. Afferma che è necessario un leader per “continuar della nostra letizia”. Sebbene stabilire un re o una regina possa sembrare una reminiscenza dei valori monarchici, il fatto che il potere venga trasferito ogni giornata dopo la canzone suggerisce valori più secolari di democrazia. La brigata ha lo scopo di onorare e obbedire al sovrano, il che assomiglia a come verrebbe trattato un monarca. Tuttavia il sovrano sta aiutando a continuare l'allegria e la gioia per l'intero gruppo. Il sovrano sembra ricoprire il ruolo di facilitatore piuttosto che di vero e proprio re o regina. Pampinea aiuta a stabilire la chiara struttura narrativa e l'ordine del *Decameron*. La scelta del

---

<sup>34</sup> Boccaccio and Branca, *Decameron*.

cantante della ballata è anche il primo atto del nuovo sovrano del giornata. Questo fatto rivela il ruolo stabilizzante delle ballate: stabiliscono il senso dell'ordine per il regno di ogni sovrano. Attraverso la stilometria e l'elaborazione del linguaggio naturale si possono stabilire la psiche emotiva della brigata e l'ordine che le ballate portano al romanzo.



## Stilometria (Psiche Emotiva)

### *Dendrogramma: Visualizzazioni della Stilometria*

Un dendrogramma è un diagramma ramificato che viene utilizzato per visualizzare somiglianze tra diversi testi letterari. In questo studio, questo metodo di visualizzazione verrà utilizzato per mostrare le somiglianze nello stile di scrittura tra le dieci ballate del *Decamerone*. Esiste una metodologia per interpretare questi diagrammi. Ogni ramo separato è chiamato clade. La fine di ogni clade è chiamata foglia. Il posizionamento di ogni clade può rivelare quanto siano simili i diversi testi. Maggiore è l'altezza tra due rami, maggiore è la differenza. Durante la lettura dall'alto verso il basso, è possibile identificare grandi gruppi. Quando si legge dal basso verso l'alto, le somiglianze tra le singole foglie possono essere notate più chiaramente. L'altezza delle linee verticali che collegano i rami indica il livello di somiglianza tra i diversi rami. Le distanze orizzontali non hanno significato in un dendrogramma. Per creare i dendrogrammi per questo studio, è stata seguita la seguente procedura:

1. Un file di testo è stato creato per ciascuna delle ballate e pulito in modo appropriato
2. La frequenza di ogni parola in ogni file di testo è stata calcolata e tabulata
3. È stata calcolata la differenza tra tutti i file di testo della ballata
4. Il dendrogramma è stato creato

I dendrogrammi riveleranno somiglianze tra le diverse ballate e costituiranno una nuova lente interpretiva per l'analisi delle ballate.

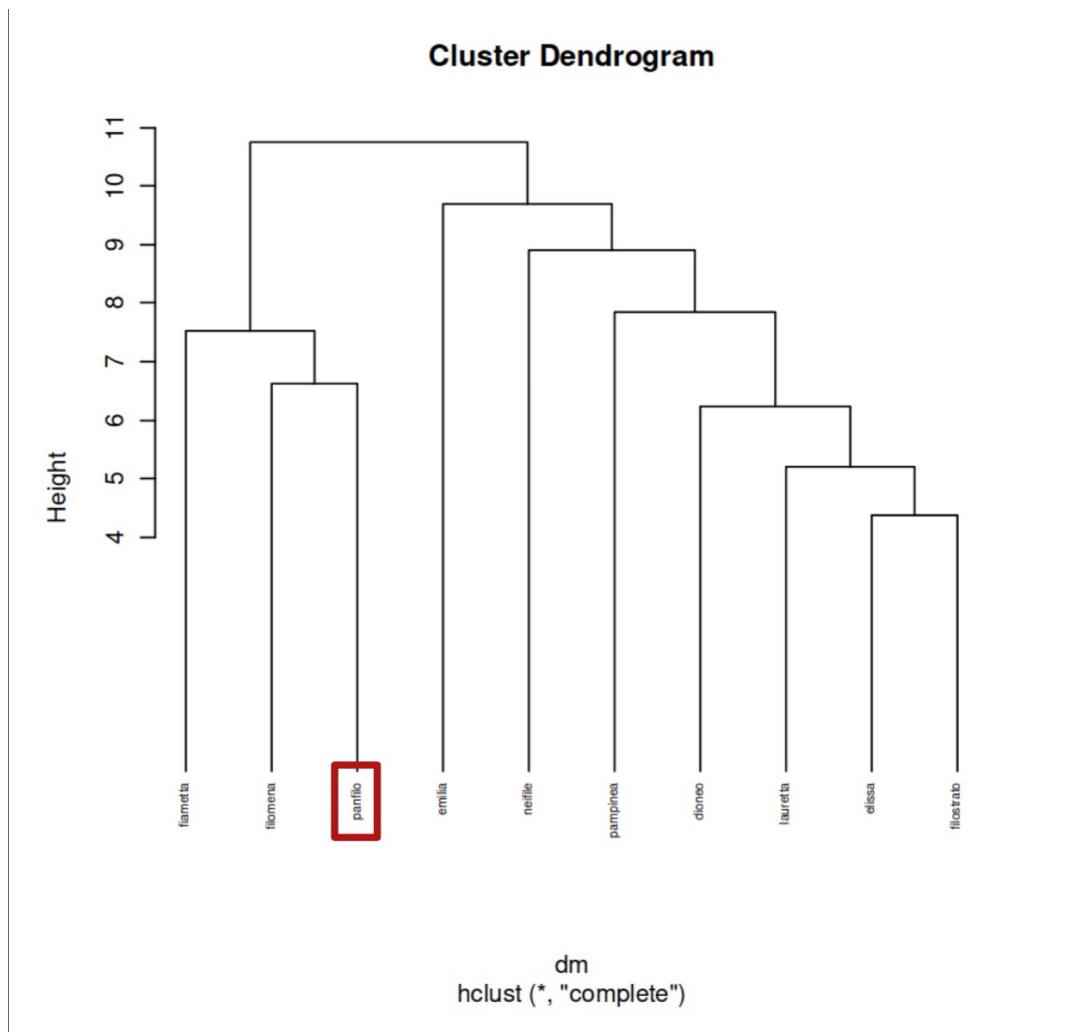
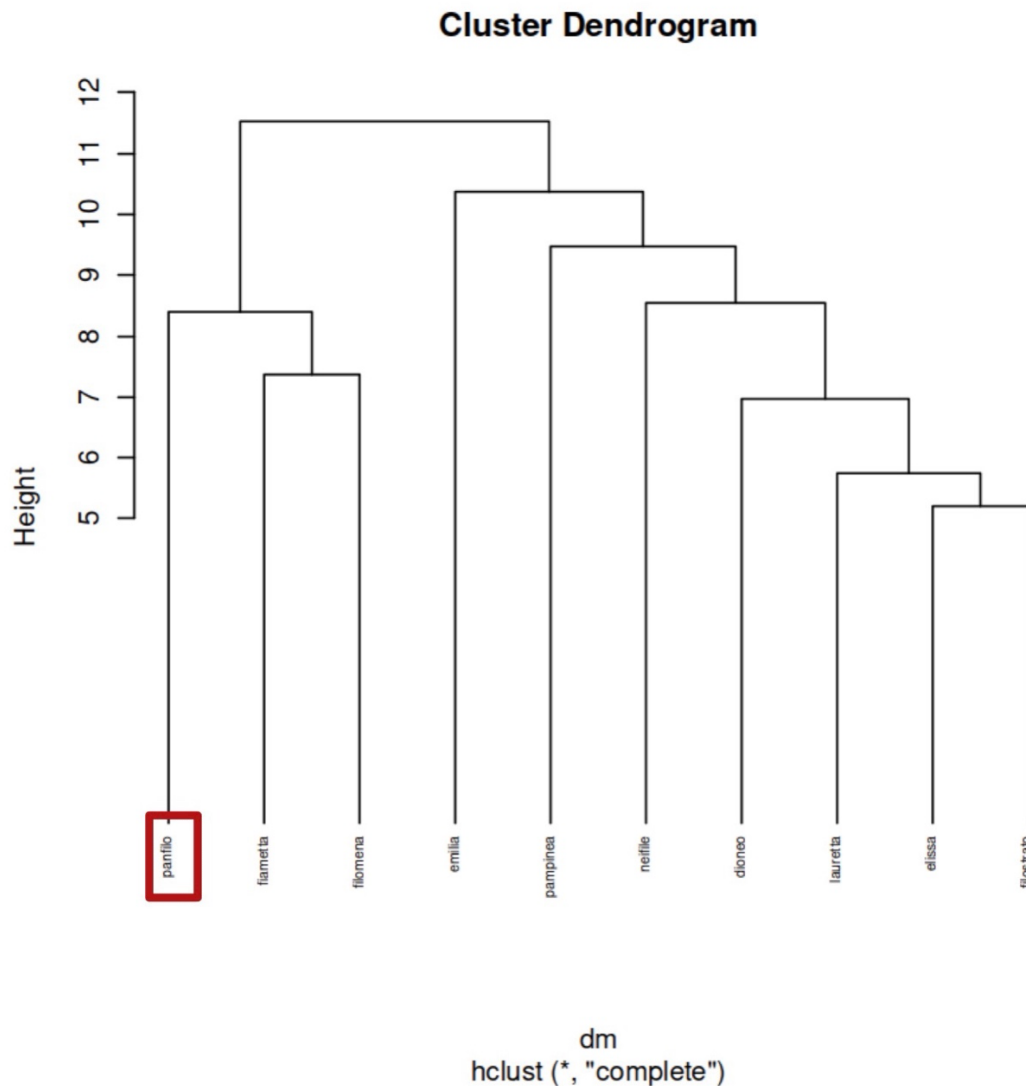
*Resultati Stilometrici*

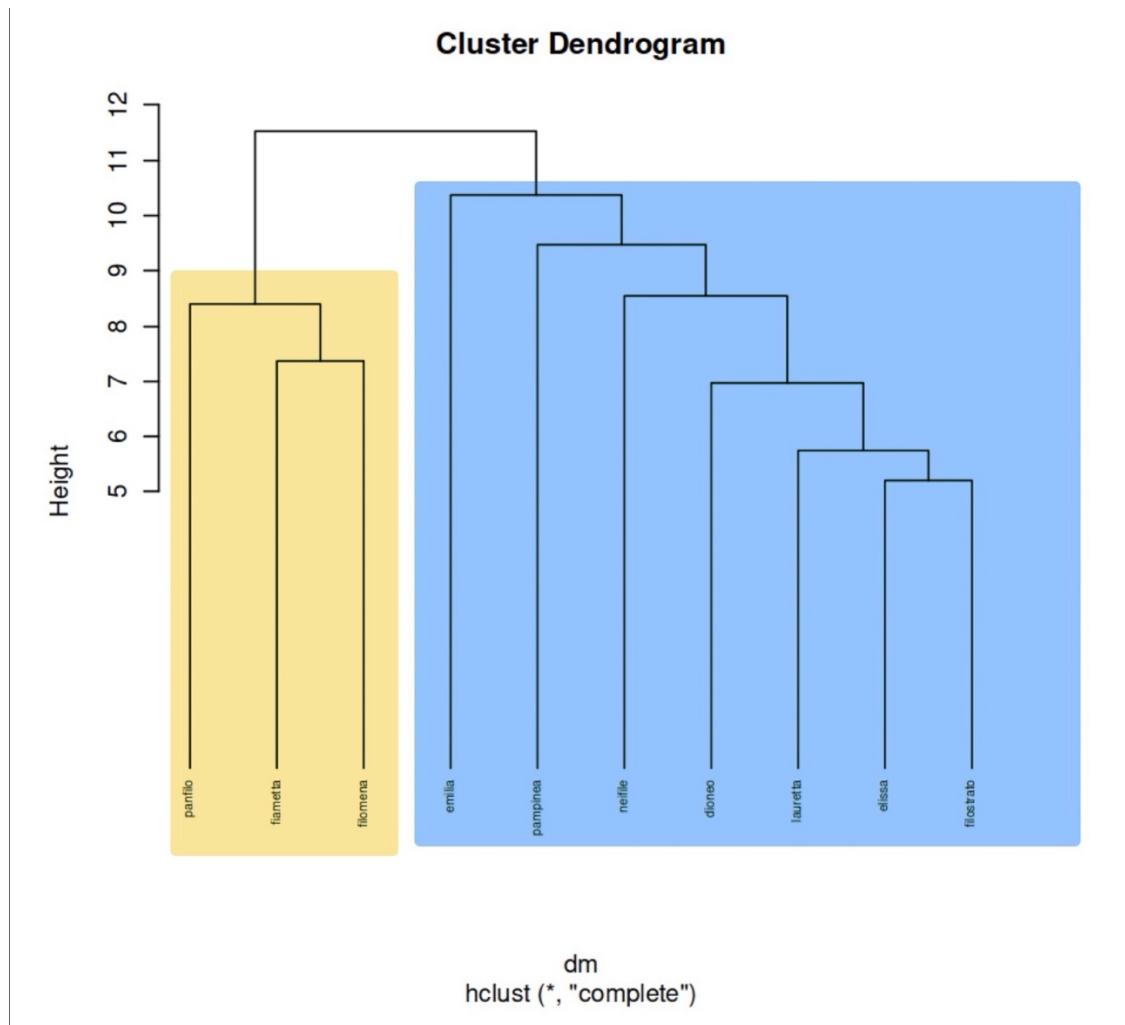
Figura 1. Dendrogramma delle dieci ballate con stop word incluse



*Figura 2. Dendrogramma delle dieci ballate con stop word escluse*

Prima di iniziare un'analisi complessa del dendrogramma, è importante notare l'impatto che gli stop word hanno sui risultati dei test stilometrici. Quando ci sono stop word, Panfilo risulta esser come più simile a Filomena; tuttavia, senza stop word, Panfilo è meno simile a Fiammetta e Filomena. Questa differenza rafforza l'importanza di rimuovere le stop word sovrarappresentate per i test che coinvolgono la frequenza delle parole. La creazione di un dendrogramma senza stop word e uno con stop word aumenta anche la fiducia del lettore nella mia scelta di rimuovere

le parole e rafforza la validità del test. Qualsiasi analisi futura non è influenzata da parole non significative sovra rappresentate e la visualizzazione di entrambi i grafici mostra la differenza.



*Figura 3. Dendrogram delle dieci ballate con due gruppi distinti. Gruppo 1 in giallo: Panfilo, Fiammetta, Filomena; Gruppo 2 in blu: Emilia, Pampinea, Neifile, Dioneo, Lauretta, Elissa, Filostrato.*

I membri della brigata sono divisi in due grandi gruppi dall'analisi stilometrica. Analizzando il diagramma dall'alto verso il basso, si distinguono due cladi principali: il gruppo 1, contrassegnato in giallo, è costituito da Panfilo, Fiammetta e Filomena e il gruppo 2, contrassegnato in blu, è costituito da Emilia, Pampinea, Neifile, Dioneo, Lauretta, Elissa,

Filostrato. Il gruppo 1 può essere ulteriormente suddiviso in due sottogruppi: un gruppo che include solo Panfilo e un altro che include Fiammetta e Filomena. Anche, il gruppo 2 può essere suddiviso in sottogruppi: Elissa e Filostrato in un gruppo ed Emilia, Pamapinea, Neifile, Dioneo e Lairetta nei propri gruppi individuali. I membri della brigata del gruppo 1 sono più simili tra loro rispetto ai membri della brigata del gruppo 2.

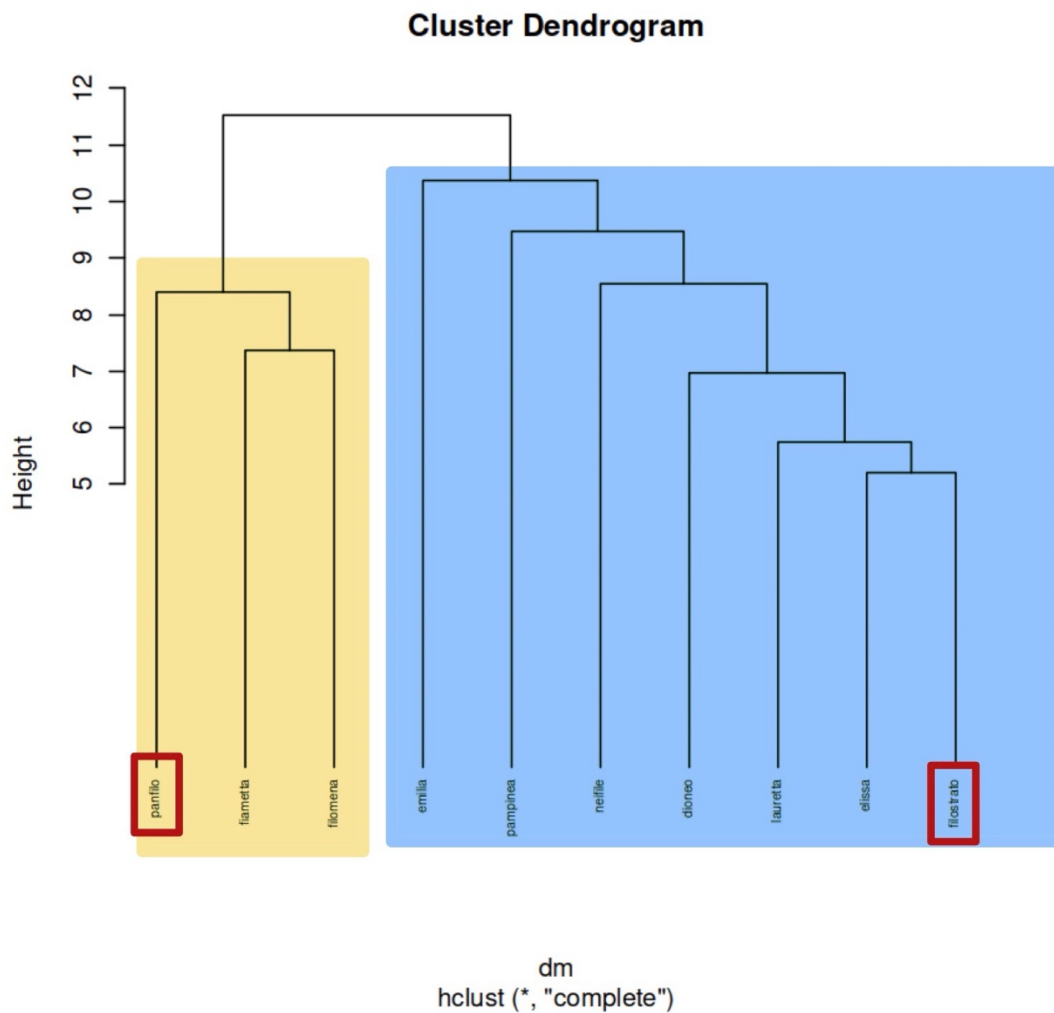


Figura 4. Dendrogramma di dieci ballate. Panfilo and Filostrato, i due uomini più vicino per età, hanno la più grande differenza nei loro stili.

Guardando il dendrogramma dal basso verso l'alto, nel Gruppo 1, le ballate di Fiammetta e Filomena sono le più simili tra loro. Nel Gruppo 2, le ballate di Filostrato ed Elissa sono molto simili. Ci si aspetterebbe che Panfilo e Filostrato, i due maschi del gruppo più vicini per età, avessero ballate simili. Invece, Panfilo e Filostrato hanno le ballate più distinte sul dendrogramma. Ancora più interessante, Panfilo è nel Gruppo 1 con donne più giovani di lui. Cosa può rivelare l'analisi delle ballate del Gruppo 1 e della ballata di Filostrato su come la psiche emotiva trascenda il genere? In che modo Boccaccio usa queste ballate per deviare dalle norme di genere previste? L'analisi delle ballate attraverso il dendrogramma mirerà a rispondere a queste domande.

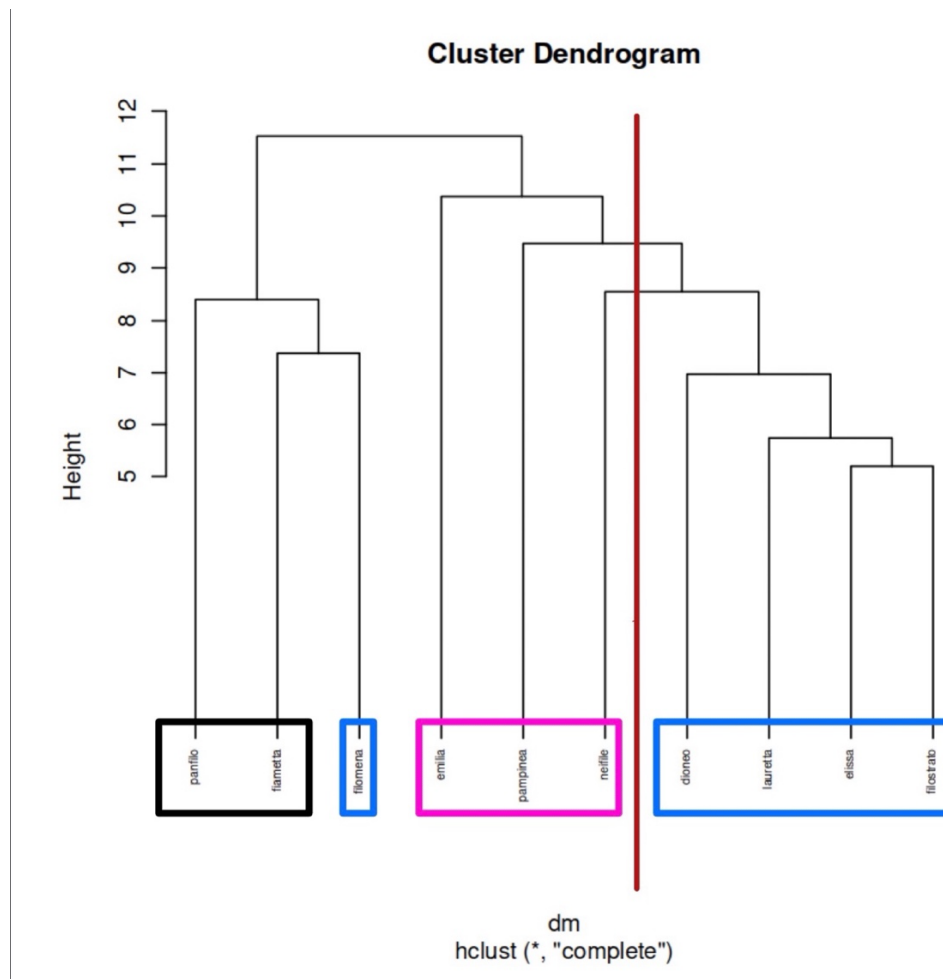


Figura 5. Dendrogramma delle dieci ballate. Il gruppo 2 è molto chiaramente diviso per tono. Dino Cervigni, un eminente studioso di letteratura italiana medievale, ha classificato ogni autore per tono in *Fiammetta's Song of Jealousy: Are The Young People Still At Play?*. Quelli in rosa hanno un tono complessivamente allegro e i nomi etichettati in blu hanno toni malinconici. Quelli in nero non sono stati definiti da Cervigni. La linea rossa divide il Gruppo 2 per tono.

Uno sguardo più attento al Gruppo 2 rivela che il gruppo è diviso in ballate con toni più tristi e in ballate con toni più allegri. Queste classificazioni tristi e felici si basano sul *Fiammetta's Song of Jealousy: Are The Young People Still At Play?* di Cervigni: i giovani sono ancora in gioco? La classificazione delle ballate di Cervigni è accurata? Quali parole guidano la scissione tra i due gruppi?

*La Ballata di Panfilo*

9 Tanto è, Amore, il benech'io per te sento,  
 e l'allegrezza e 'l gioco, ch'io son felice  
 ardendo nel tuo foco.

10 L'abondante allegrezza ch'è nel core,  
 dell'alta gioia e caranella qual m'hai recato,  
 non potendo capervi esce di fore,  
 e nella faccia chiaramostra 'l mio lieto stato;  
 ch'essendo innamorato  
 in così alto e raguardevol loco  
 lieve mi fa lo star dov'io mi coco.

11 Io non so col mio canto dimostrare,  
 né disegnar col dito, Amore, il ben ch'io sento;  
 e s'io sapessi, mel convien celare;  
 ché, s'el fosse sentito,  
 torneria in tormento:  
 ma io son sí contento,  
 ch'ogni parlar sarebbe corto e fioco  
 pria n'avessi mostrato pure un poco.

12 Chi potrebbe estimar che le mie braccia  
 aggiugnesser già mailà dov'io l'ho tenute,  
 e ch'io dovessi giunger la mia faccia  
 là dov'io l'accostai  
 per grazia e per salute?  
 Non mi sarien credute  
 le mie fortune; ond'io tutto m'infoco,



*quel nascondendo ond'io m'allegro e gioco.*<sup>35</sup> (VIII, Conclusione, 9-12)

La ballata di Panfilo esplora il tema dell'amore e la gioia e la felicità che porta. La ballata dimostra anche il dubbio che un amore nascosto porta a una persona.

Nel primo verso, Panfilo esprime il proprio amore per qualcuno e la felicità e la gioia che prova quando è con quella persona. Usa la metafora del fuoco per descrivere la sua passione e l'intensità dei suoi sentimenti. Panfilo è così felice e gioioso che questa felicità è evidente nel suo viso e nel linguaggio del corpo. L'uso delle parole "ardendo" e "foco" suggerisce che l'amore di chi parla è appassionato e intenso, e per questo si sente vivo. L'Amore intenso come un fuoco sembra una cosa positiva ma dà anche cicatrici a coloro che sono coinvolti. Inoltre, "foco" suggerisce che Panfilo proverà dolore, così come per una malattia d'amore. Panfilo identifica la persona di cui è innamorato come 'Amore' svelando che lui ama così tanto questa persona che lei è diventata la personificazione dell'amore nella sua vita. Questa persona lo porta a uno stato di "bene." L'uso delle parole "allegrezza" e "gioco" mostra che l'amore che Panfilo ha per questa persona gli porta vera gioia ed è qualcosa da celebrare.

La seconda strofa continua con il tema della gioia e della felicità, mentre la cantante parla dell'abbondante gioia nel suo cuore che la persona che ama gli ha portato : "L'abbondante allegrezza ch'è nel core". La scelta della parola 'abbondante' suggerisce l'esistenza di un amore scarso. Se l'amore è abbondante ora, è stato scarso in passato? La cantante suggerisce che la sua gioia è così grande che non può essere contenuta dentro di lui e si riversa sul suo viso, rivelando la sua felicità al mondo. La sua faccia ora è 'chiara', ma ancora una volta Panfilo accenna ad

---

<sup>35</sup> Boccaccio and Branca.

ombre passate. Questo amore lo ha fatto piangere in passato? L'uso della parola "caranella" suggerisce che la persona amata da chi parla è molto importante e che ha a cuore la loro relazione. La frase "non potendo capervi esce di fore" enfatizza che altre persone non possono capire l'estensione dell'amore di Panfilo. Inoltre, questa frase allude all'isolamento che l'intensità di questo amore porta a Panfilo.

La terza strofa mostra la difficoltà di chi parla nell'esprimere il proprio amore e la gioia che ne deriva: 'Io non so col mio canto dimostrare, né disegnar col dito, Amore, il ben ch'io sento'. Panfilo non ha la capacità di trasmettere adeguatamente il suo amore attraverso l'arte. È in un territorio inesplorato. Sebbene affermi ripetutamente la sua felicità, la sua incapacità di esprimersi la mette in dubbio. La confusione che il suo amore porta crea dissonanza e tensione tra la sua psiche interiore e l'espressione esteriore. Sebbene questo amore lo faccia sentire bene, la sua incapacità di comprenderlo intellettualmente è molto probabilmente frustrante per questo giovane. La cantata suggerisce che la sua felicità è così grande che è impossibile trasmetterla completamente a parole. L'uso della parola "celare" suggerisce che chi parla potrebbe essere riluttante a esprimere i propri sentimenti perché teme che possa causargli dolore se viene rifiutato, entrando in uno stato di 'tormento'. Ama ma il suo amore lo rende vulnerabile al rifiuto.

La strofa finale suggerisce che l'amore di chi parla è così grande che è difficile per gli altri capire o credere. La cantata suggerisce che la loro fortuna è così grande che non riesce nemmeno a crederci: 'Non mi sarien credutele mie fortune'. Anche Panfilo non crede quanto sia fortunato ad essere innamorato. L'uso della parola "infoco" suggerisce che l'amore di chi parla sta ancora bruciando forte e ne è consumato. Il verso finale suggerisce che Panfilo potrebbe tenere nascosta

la propria gioia e felicità al mondo, forse per paura che possa essergli portata via. Perché è consapevole della sua capacità di essere amato dagli altri? Ha paura che il suo amore gli venga portato via dal mondo? Forse la sua insicurezza deriva dalla possibilità che la peste uccida la sua amata. Perché rivelare l'amore al mondo se potrebbe essere portato via?

Nel complesso, la ballata esplora il tema dell'amore e la gioia e la felicità che porta. La passione e l'intensità della cantata sono evidenti in tutta la ballata e l'uso della metafora del fuoco e delle immagini aiuta a trasmettere la profondità dei suoi sentimenti. Panfilo esprime anche i suoi timori e dubbi nei confronti del suo luminoso Amore. Alla fine, la sua confusione si rispecchia nella risposta degli altri membri della brigata.

La canzone di Panfilo aveva fine, alla quale quantunque per tutti fosse compiutamente risposto, niun ve n'ebbe che, con più attenta sollecitudine che a lui non apparteneva, non notasse le parole di quella, ingegnandosi di quello volersi indovinare che egli di convenirgli tener nascoso cantava; e quantunque varii varie cose andassero imaginando, niun per ciò alla verità del fatto pervenne. Ma la reina, poi che vide la canzon di Panfilo finita e le giovani donne e gli uomini volentier riposarsi, comandò che ciascuno se n'andasse a dormire (VIII, Conclusione, 13).<sup>36</sup>

Gli altri membri della brigata sono in grado di cogliere uno strato 'nascoso' nella sua canzone. Tuttavia, nessuno è stato in grado di indovinare la 'verità'. La conversazione finisce quando la regina del giornata manda tutti a dormire. Proprio come Panfilo, tutti devono dormire su emozioni vorticose. Alla fine dimenticheranno, ma Panfilo dovrà sopportare il dolore del destino sconosciuto del suo ardente amore.

---

<sup>36</sup> Boccaccio and Branca.

*La Ballata di Filostrato*

*11 Lagrimando dimostro  
quanto si dolga con ragione il core  
d'esser tradito sotto fede, Amore.*

*12 Amore, allora che primieramente  
ponesti in lui colei per cui sospiro,  
senza sperar salute,  
sí piena la mostrasti di virtute,  
che lieve reputava ogni martiro,  
che per te nella mente,  
ch'è rimasa dolente,  
fosse venuto; ma il mio errore  
ora conosco, e non senza dolore.*

*13 Fatto m'ha conoscente dello 'nganno  
vedermi abbandonato da colei,  
in cui sola sperava;  
ch'allora ch'i'  
piú esser mi pensava  
nella sua grazia e servidore a lei,  
senza mirare al danno  
del mio futuro affanno,  
m'accorsi lei aver l'altrui valore  
dentro raccolto, e me cacciato fore.*

*14 Com'io conobbi me di fuor cacciato,  
nacque nel core un pianto doloroso,  
che ancor vi dimora:*

*e spesso maladico il giornata e l'ora  
 che pria m'apparve il suo viso amoroso  
 d'alta biltate ornato  
 e piú che mai 'nfiammato!  
 La fede mia, la speranza e l'ardore  
 va bestemmiando l'anima che more.*

*15 Quanto 'l mio duol senza conforto sia,  
 signor, tu ' puoi sentir, tanto ti chiamo  
 con dolorosa voce:  
 e dicoti che tanto e sí mi cuoce,  
 che per minor martir la morte bramo.  
 Venga dunque, e la mia  
 vita crudele e ria  
 termini col suo colpo, e 'l mio furore,  
 ch'ove ch'io vada il sentirò minore.*

*16 Null'altra via, niuno altro conforto  
 mi resta piú che morte alla mia doglia.  
 Dallami dunque omai;  
 pon fine, Amor, con essa alli miei guai,  
 e 'l cor di vita sí misera spoglia.  
 Deh fallo, poi ch'a torto  
 m'è gioia tolta e diporto.  
 Fa' costei lieta, morend'io, signore,  
 come l'hai fatta di nuovo amadore.*

*17 Ballata mia, se alcun non t'apparia  
 non men curo, per ciò che nessuno,  
 com'io, ti può cantare.  
 Una fatica sola ti vo' dare:*

*che tu ritruovi Amore, e a lui solo uno  
 quanto mi sia discara  
 la trista vita amara  
 dimostri appien, pregandol che 'n migliore  
 porto ne ponga per lo suo onore<sup>37</sup> (IV, Conclusione, 11-17).*

La ballata di Filostrato è una ballata che esprime il dolore e la sofferenza di un uomo che è stato tradito dalla donna che ama. Si castiga per non aver avuto la lungimiranza di prevedere le trasgressioni della sua amante, desidera che il Signore lo uccida e spera che nessuno debba mai condividere il suo dolore. La ballata utilizza una serie di dispositivi letterari per trasmettere le emozioni e i pensieri di chi parla.

La prima e la seconda strofa introducono il tema della ballata, che è l'amore di chi parla per una donna che lo ha tradito 'Lagrimando dimostro quanto si dolga con ragione il core d'esser tradito sotto fede'. 'Lagrimando' stabilisce chiaramente che Filostrato ha pianto. Il suo dolore non è nascosto. 'Esser tradito sotto fede' mostra che Filostrato era stato tradito dalla sua innamorata in cui lui aveva fede. Si rivolge ad Amore, la personificazione dell'amore, e riflette su come inizialmente credeva che la donna che amava 'sí piena la mostrasti di virtute' e non gli avrebbe mai fatto del male. Tuttavia, da allora ha saputo che lei lo ha ingannato e ora è pieno di dolore e rimpianto. Invece di incolpare lei, dà la colpa a se stesso: 'ma il mio errore, ora conosco'. Si assuma la colpa per tutto il suo dolore.

La terza strofa rivela l'entità dell'angoscia di chi parla, quando si rende conto che la donna che amava lo ha abbandonato per qualcun altro. È pieno di amarezza e risentimento mentre ricorda

---

<sup>37</sup> Boccaccio and Branca.

come si era dedicato a lei, solo per essere messo da parte quando ha trovato qualcuno di migliore. L'uso da parte della cantante della parola "nganno" sottolinea il tradimento che sente, e l'uso del verbo "cacciato" sottolinea il suo senso di rifiuto ed espulsione dalla vita della donna. Il tradimento del suo amante non lo rendeva solo triste ma lo ha spinto completamente fuori dalla sua vita, un posto in cui aveva sperato di trascorrere tutta la sua vita. Lontano da lei, si sente 'abbandonato'. Ha riposto le sue speranze in lei e il suo amore è stato gettato via.

Nella quarta strofa, il dolore della cantante si intensifica mentre riflette sui suoi sentimenti di perdita e disperazione: 'spesso maladico il giornata e l'ora che pria m'apparve il suo viso amoroso.' La parola 'spesso' rivela la sua ossessione per il suo dolore. È consumato dal dolore e piange per l'amore che ha perso. Maledice il suo destino e desidera la morte come liberazione dalla sua sofferenza. L'uso da parte della cantante delle parole "bestemmiando" e "anima che more" trasmette il suo senso di disperazione.

Nella quinta strofa, la cantante supplica Dio di porre fine alla sua miseria e porre fine alla sua vita. Vede la morte come l'unica via di fuga dal dolore che prova e implora Amore di mettere fine alla sua sofferenza. Il suo uso della parola "furore" sottolinea il suo senso di essere spinto sull'orlo della follia dalla sua sofferenza. Le sue emozioni non hanno virgiola. È completamente in bìa dell'angoscia e della disperazione. Non vedendo alcuna via per tornare all'equilibrio e alla speranza, implora Dio di porre fine la sua miseria attraverso la morte. Crede che solo nella morte le sue emozioni saranno minori. L'uso della parola 'minore' descrive la diminuzione della gravità delle sue emozioni e l'innocenza che gli sarà restituita senza questo dolore. Potrà sentire com'era quando era più giovane e portare ancora l'innocenza di chi non è stato ancora tradito.

La sesta strofa sottolinea il senso di disperazione di chi parla, poiché non vede altra via d'uscita dalla sua miseria che la morte. Sembra chiedere ad Amore e al Signore di porre fine alla sua vita. Non è chiaro se stia confondendo o meno il suo Amore con Dio. Poiché la infedeltà della sua donna lo ha colpito così tanto, lei può essere vista come un dio che governa la sua vita. Chiede al Signore di porre fine alla sua sofferenza e di rendere felice la donna che lo ha tradito, anche se questo significa la sua stessa morte. L'uso da parte della cantante della frase "vita crudele" sottolinea la sua sensazione di essere intrappolato in una vita insopportabile. Questa frase 'fa' costei lieta, morend'io, signore, come l'hai fatta di nuovo amadore' ricorda al lettore che Filostrato si incolpa della sua sofferenza e della sua infedeltà. Vede la fine della sua vita come un mezzo per porre fine alla sua sofferenza e permettere alla sua amante di essere felice.

La strofa finale è una ballata, in cui Filostrato invita la ballata stessa a testimoniare il suo dolore e a chiedere ad Amore di trovare qualcuno che possa amarlo come merita. Chiede alla ballata di trovare Amore e di supplicarlo di portargli un po' di sollievo dalla sua miseria. Il verso finale, "porto ne ponga per lo suo onore", sottolinea il senso di essere stato offeso e il suo desiderio di giustizia. In questa strofa, Filostrato riesce ad incolpare qualcosa di diverso da se stesso.

Nel complesso, la ballata di Filostrato è una potente espressione del dolore e della sofferenza che possono derivare dal tradimento in amore. L'uso del linguaggio e delle immagini da parte di chi parla trasmette chiaramente il suo senso di perdita e disperazione, e la sua richiesta di sollievo dalla sua miseria è toccante e commovente. L'uso della forma della ballata aumenta il suo impatto emotivo e i suoi temi di amore, tradimento e perdita sono senza tempo e universali.



*Panfilo e Filostrato: Psiche Emotiva Trascende il Genere*

L'analisi stilometrica ha rivelato che Filostrato e Panfilo, i due uomini più vicini per età, hanno le ballate più distinte di tutti i membri della brigata. Questa dissonanza riflette come il sesso e l'età non determinano la psiche emotiva di una persona. Invece, l'esperienza guida le emozioni e questo fenomeno può essere rivelato attraverso la ballata, una forma che Boccaccio afferma essere come un percorso verso l'anima. Le due ballate sono piuttosto diverse sia nel contenuto che nello stile. La ballata di Panfilo è una ballata d'amore che esprime i sentimenti di gioia e felicità di chi parla nell'essere innamorato e anche la confusione che deriva dal provare amore ardente per la prima volta. La ballata di Filostrato, invece, è un lamento per un amore perduto ed esprime il profondo dolore e la disperazione di chi parla. La ballata di Panfilo è scritta in uno stile più leggero e giocoso, mentre la seconda è più seria e dolente.

Nella ballata di Panfilo, la cantante esprime il suo amore per qualcuno, un partner romantico, e quanta gioia e felicità prova a causa di quell'amore. Dice che non può esprimere appieno la profondità dei suoi sentimenti attraverso parole o gesti, ma che la sua felicità si vede sul suo volto. Implica anche che il suo amore per questa persona lo ha elevato a un livello superiore, facendolo sentire più leggero e spensierato. Il tono della ballata è molto positivo e ottimista, con la cantante che si crogiola nella gioia del suo amore. La ballata porta anche mancanza di chiarezza su ciò che sta accadendo esattamente nel suo amore. Panfilo non è in grado di esprimere chiaramente a parole come si sente.

Nella ballata di Filostrato, la cantante esprime il suo profondo dolore e la sua disperazione per essere stato tradito da qualcuno che amava. Si lamenta del fatto di aver riposto così tanta fede e

fiducia in questa persona, solo per essere abbandonato e messo da parte. Esprime il suo dolore e la sua rabbia per la situazione, maledicendo il giornata in cui ha incontrato questa persona. Il tono della ballata è molto più cupo e triste del primo, con la cantate consumato dal dolore e dalla disperazione. A differenza di Panfilo, Filostrato esprime sempre chiaramente su ciò che prova. Usa un linguaggio estremamente forte per rappresentare chiaramente il suo stato cupo.

Le due ballate differiscono anche nell'uso del linguaggio e delle tecniche poetiche. La prima utilizza molte immagini e un linguaggio figurativo per trasmettere le emozioni di chi parla, come "bruciare nel tuo fuoco" e "gioia traboccante". Utilizza anche molte ripetizioni, con la frase "il bene ch'io per te sento" ripetuta più volte durante il poema. La ballata di Filostrato utilizza una lingua più semplice e meno tecniche poetiche, affidandosi maggiormente all'emozione cruda di chi parla per trasmettere il suo messaggio.

Nel complesso, le due ballate rappresentano atteggiamenti molto diversi nei confronti dell'amore e delle relazioni. La ballata di Panfilo celebra il potere di trasformazione dell'amore, mentre la ballata di Filostrato ritrae il dolore e il crepacuore che possono venire quando quell'amore viene tradito. Entrambi i poeti usano il linguaggio e le immagini per evocare forti emozioni nel lettore, e la struttura delle loro ballate fornisce un quadro per esplorare questi sentimenti complessi e spesso contrastanti. Sono entrambi potenti esempi del modo in cui la ballata può catturare la complessità e la profondità delle emozioni umane. Attraverso questi due personaggi, Boccaccio mostra abilmente come l'esperienza guidi la psiche emotiva e non il genere.

*Raggrupparsi con Fiammetta e Filomena Fa Chiarezza sulle Emozioni di Panfilo*

Mentre la ballata di Fiammetta si concentra sulle complessità dell'amore e della gelosia, la ballata di Panfilo presenta un tono più celebrativo nei confronti dell'amore. La ballata di Panfilo esprime pura gioia e contentezza per l'essere innamorati, mentre la ballata di Fiammetta mette in mostra l'ansia e le paure che derivano dall'amore, soprattutto per quanto riguarda la gelosia.

L'accostamento di Panfilo con Fiammetta nel test stilometrico aggiunge chiarezza al motivo per cui Panfilo non riesce ad esprimersi. Come Fiammetta, Panfilo ha ansia e paura che circondano l'intensità del suo amore ardente.

La ballata di Panfilo riguarda anche più l'aspetto fisico dell'amore, con descrizioni di abbracci e contatto fisico. La ballata di Fiammetta, invece, sottolinea il tumulto emotivo e mentale che l'amore può portare, in particolare quando si tratta di preoccuparsi della fedeltà del proprio amante.

Nonostante queste differenze, entrambe le ballate sono unite nella loro esplorazione del potere e della complessità dell'amore. Entrambi i poeti mostrano la natura trasformativa dell'amore, come può ispirare le persone a diventare versioni migliori di se stesse e ad abbracciare la vita pienamente. Inoltre, entrambi i poeti usano i loro versi per esplorare il ruolo delle norme sociali e delle aspettative in amore. La ballata di Fiammetta evidenzia l'aspettativa della società per le donne di essere miti e obbedienti, in particolare quando si tratta del loro rapporto con i loro partner. La ballata di Panfilo, invece, sfida la norma sociale secondo cui l'amore è fonte di dolore e sofferenza, celebrando invece la felicità e la gioia che l'amore può portare.

Le ballate di Panfilo e Filomena sono raggruppate perché accomunate da temi simili. Sia la ballata di Filomena che quella di Panfilo trattano dell'esperienza dell'innamoramento e delle intense emozioni che ne derivano. La ballata di Filomena esprime il suo desiderio di tornare nel luogo in cui si è innamorata e il disperato desiderio di ricongiungersi con l'amato. Descrive il suo amore per lui come un desiderio ardente che la consuma e non riesce a trovare pace finché non si riunisce al suo amante. Invece, la ballata di Panfilo esprime la gioia e la felicità che lui prova per il suo amore per questa persona. Descrive la sua felicità come traboccante ed è incapace di contenerla.

In termini di tono, la ballata di Filomena ha un tono più malinconico e nostalgico, mentre la ballata di Panfilo ha un tono più gioioso e celebrativo. L'uso di Filomena di frasi come "noiosa gioco" e "virtù disgottita" suggerisce un senso di tristezza e disperazione, mentre l'uso di frasi come "abbondante gioia" e "felice ardere nel tuo fuoco" da parte di Panfilo suggerisce un senso di gioia e celebrazione. In termini di linguaggio, entrambe le ballate usano un linguaggio ricco e descrittivo per trasmettere le loro emozioni. La ballata di Filomena usa frasi come "disio ficoso" e "anima persosta", mentre la ballata di Panfilo usa frasi come "l'allegrezza e 'l gioco" e "caranella".

Nel complesso, sebbene entrambe le ballate trattino il tema dell'amore, differiscono per tono e prospettiva. La ballata di Filomena esprime il dolore e il desiderio associati all'amore, mentre la ballata di Panfilo esprime gioia e felicità. Mentre le ballate di Fiammetta e Panfilo possono differire per tono e contenuto, entrambe le opere alla fine esplorano le molte sfaccettature dell'amore, evidenziandone la natura trasformativa e complessa. La collocazione di Panfilo insieme a Fiammetta e Filomena aggiunge chiarezza alla confusione che lui e gli altri membri

della brigata provano nei confronti della sua ballata. Inoltre, il posizionamento della ballata di Panfilo con due donne, una delle quali è significativamente più anziana di lui (Filomena), rivela che psiche emotiva non segue semplici linee di genere. Invece, anche nel trecento, ogni individuo sviluppa una psiche emotiva che è guidata dalle proprie esperienze e dalla propria risposta unica a quell'esperimento.

### *Gruppo 2: La Linea tra Felicità e Tristezza*

La felicità e la tristezza sono emozioni umane complesse che compaiono nelle dieci ballate del *Decameron*. Nel gruppo 2 del dendrogramma, i membri della brigata sono divisi dai loro toni, con felice a sinistra e triste a destra. Le classificazioni sono state stabilite da Dino Cervigni, rispettato studioso di Boccaccio, in “Fiammetta's Song of Jealousy Are The Young People”. Per analizzare l'esattezza di questa scissione e scoprire cosa unisce i cantanti tristi e felici del Gruppo 2, verrà condotto un confronto tra le ballate di Neifile e Dioneo. Queste ballate sono state scelte perché sono le due ballate al limite dei gruppi felici e tristi. La ballata di Neifile e la ballata di Dioneo affrontano entrambi il tema dell'amore, ma differiscono per approccio e stile. Mentre Neifile e Dioneo decidono entrambe di mostrare ammirazione per i loro amanti attraverso l'apprezzamento della natura, Dioneo assume un tono molto più cupo. Il suo amore non è corrisposto e l'amore di Neifile sembra essere completamente ricambiato. La differenza nei loro stili riflette le loro esperienze personali e le loro reazioni emotive a queste esperienze.

#### *La Ballata di Dioneo*

*16 Amor, la vaga luce che move  
da' begli occhi di costei  
servo m'ha fatto di te e di lei.*

*17 Mosse da' suoi begli occhi lo splendore*

*che pria la fiamma tua nel cor m'accese,  
 per li miei trapassando;  
 e quanto fosse grande il tuo valore,  
 il bel viso di lei mi fé palese;  
 il quale imaginando,  
 mi senti' gir legando  
 ogni virtù e sottoporla a lei,  
 fatta nuova cagion de' sospir miei.*

*18 Così de' tuoi, adunque, divenuto  
 son, signor caro, e ubidiente aspetto  
 dal tuo poter merzede;  
 ma non so ben se 'ntero è conosciuto  
 l'alto disio che messo m'hai nel petto  
 nè la mia intera fede  
 da costei, che possiede  
 sí la mia mente, che io non torrei  
 pace fuor che da essa, né vorrei.*

*19 Per ch'io ti priego, dolce signor mio,  
 che gliel dimostri e faccile sentire  
 alquanto del tuo foco  
 in servizio di me, ché vedi ch'io  
 già mi consumo amando e nel martire  
 mi sfaccio a poco a poco;  
 e poi, quando fia loco,  
 me raccomanda a lei, come tu dei,  
 che teco a farlo volentier verrei.<sup>38</sup> (V, Conclusione, 16-19)*

---

<sup>38</sup> Boccaccio and Branca.

*La Ballata di Neifile*

8 *Io mi son giovinetta, e volentieri  
m'allegro e canto en la stagion novella,  
merzé d'amore e de' dolci pensieri.*

9 *Io vo pe' verdi prati riguardando  
i bianchi fiori e' gialli e i vermigli,  
le rose in su le spine e' bianchi gigli,  
e tutti quanti gli vo somigliando  
al viso di colui che me amando  
ha presa e terrà sempre, come quello  
ch'altro non ha in disio che' suoi piaceri.*

10 *De' quali quand'io ne truovo alcun che sia,  
al mio parer, ben simile di lui,  
il colgo e bacio e parlo mi con lui:  
e com'io so, così l'anima mia  
tutta gli apro e ciò che 'l cor disia:  
quindi con altri il metto in ghirlandella  
legato co' miei crin biondi e leggieri.*

11 *E quel piacer che di natura il fiore  
agli occhi porge, quel simil mel dona  
che s'io vedessi la propia persona  
che m'ha accesa del suo dolce amore:  
quel che mi faccia piú il suo odore  
esprimer nol potrei con la favella,  
ma i sospir ne son testimon veri.*

12 *Li quai non escon già mai del mio petto,  
come dell'altre donne, aspri né gravi,*

*ma se ne vengon fuor caldi e soavi  
 e al mio amor sen vanno nel cospetto:  
 il qual, come gli sente, a dar diletto  
 di sé a me si move e viene in quellach'i'  
 son per dir: "Deh! vien, ch'i' non disperi".<sup>39</sup> (IX, Conclusione, 8-12)*

La ballata di Neifile è il racconto in prima persona di una giovane donna che esprime i suoi sentimenti di amore e desiderio per il suo amato. Descrive come è rapita dalla bellezza della natura e come tutto le ricorda il volto del suo amante. Parla anche del piacere che le deriva dalla vista, dall'olfatto e dal toccare i fiori e di come le ricordano il suo amante: 'E quel piacer che di natura il fiore agli occhi porge. 'Confrontando il suo amore con un fiore', Neifile suggerisce che il suo amore è naturalmente bello. Questa bellezza le piace e la fa sentire speciale rispetto alle 'altre donne.' La ballata è caratterizzata da un linguaggio semplice e diretto e da un tono chiaro e lirico. La poesia di Dioneo, invece, è più complesso ed elaborato. È scritta in terza persona ed esplora il rapporto di chi parla con l'amore e l'amato. La poesia utilizza metafore e immagini per trasmettere le emozioni e le esperienze di chi parla. La cantata descrive come la bellezza degli occhi dell'amato lo abbia reso schiavo e come ora sia un servitore dell'amore: 'la vaga luce che move da' begli occhi di costei servo m'ha fatto di te e di lei.' Questa frase dimostra che Dioneo è completamente rapito dalla sua amante. Lei era la luce che brillava e guidava le sue azioni. Parla anche del conflitto tra il suo desiderio per l'amata e la sua fedeltà al suo padrone, che gli ha dato il dono dell'amore. Dioneo è sottomesso al potere dell'amore e alla donna che desidera, ma sa che non può averla. Questa tristezza è sottolineata dall'uso di parole come "martire" e "sfaccio" per descrivere il tumulto interiore di chi parla. Al contrario, la ballata di Neifile "Io mi son

---

<sup>39</sup> Boccaccio and Branca.



giovinetta" ha un tono più gioioso, in quanto celebra i piaceri dell'amore e la bellezza della natura.

Una somiglianza tra le ballate di Neifile e Dioneo è che entrambe usano la natura per descrivere la passione di chi parla per qualcun altro. Entrambe le poesie usano immagini naturali per trasmettere la bellezza dei loro amati, con Neifile che descrive i fiori e Dioneo che descrive la luce negli occhi della sua amata. Inoltre, entrambe le ballate esprimono un senso di desiderio: Neifile desidera ardentemente la presenza dell'amante e si diverte a paragonare la bellezza della natura alla propria bellezza, mentre la cantate di Dioneo esprime la sua sottomissione al potere dell'amata e il suo desiderio di stare con lei. La natura unisce tematicamente le ballate tristi e felici nel Gruppo 2.

Entrambe le ballate usano anche un linguaggio che enfatizza l'intensità delle emozioni di chi parla. Neifile usa un linguaggio appassionato e descrittivo per descrivere il suo amore e i suoi desideri, mentre la cantate di Dioneo parla della sua sofferenza e dell'intensità del suo desiderio. In particolare, Neifile dice: 'così l'anima miatututta gli apro.' Il suo uso della parola anima mostra che l'amore che prova non è solo fisico. La parola "miatututta" sottolinea quanto sia intenso il suo amore, quanto la sua intera anima si sta aprendo. Questa immagine rispecchia quasi il linguaggio che accompagna una conversione religiosa. La sua anima si sta aprendo a un nuovo amore che influenza le sue percezioni della natura. Anche, Dioneo usa un linguaggio molto intenso: "già mi consumo amando e nel martiremi sfaccio a poco a poco". La parola "mi consumo" mostra che è completamente adsorbito dalla sua infatuazione per l'amante. La frase "nel martiremi sfaccio a poco a poco" utilizza un intenso immaginario religioso per rappresentare come Dioneo si stia consumando mentre aspetta che la sua amante ricambi il suo amore. Nel

complesso, sebbene il tono e lo stile delle due ballate differiscano, entrambe esplorano temi simili di amore, natura e passione. Queste due bellissime ballate d'amore uniscono le ballate felici e tristi del Gruppo 2 attraverso stili di scrittura che usano la natura e le immagini religiose per rivelare la complessa psiche emotiva dei loro autori.

### *Riassunto*

La stilometria è stata utilizzata per rivelare come le ballate possano scoprire la psiche emotiva dei membri della brigata. In particolare, dimostrando come la psiche sia sviluppata dalle esperienze amorose di ciascun membro e non da identificatori tipici, come il sesso e l'età. Ci si aspettava che Panfilo e Filostrato, i due membri di brigata maschi più vicini per età, condividessero le ballate simili più stilistiche, ma un test stilometrico ha rivelato che era vero il contrario. Invece Panfilo è stato raggruppato con Fiammetta e Filomena, due donne che lottano per superare la gelosia e le relazioni a distanza. Questo raggruppamento ha dato chiarezza alla natura ambigua della sua poesia. Pur vivendo un amore intenso, la novità del rapporto lo riempie di dubbi e la paura di perdere l'amore lo rattrista. In Gruppo 2, un test stilometrico ha confermato le somiglianze nei toni tristi e felici. Dividere correttamente i membri della brigata suggerisce che l'algoritmo sta riconoscendo con successo lo stile e dà valore alle affermazioni fatte su Panfilo.

## Elaborazione del Linguaggio Naturale (Ordine)

L'elaborazione del linguaggio naturale (ELN) è un sottocampo interdisciplinare di linguistica, informatica e intelligenza artificiale che si concentra sulle interazioni tra computer e linguaggio umano, in particolare su come programmare i computer per elaborare e analizzare grandi quantità di dati del linguaggio naturale. Sarà utilizzato per prevedere l'autore di ciascuna delle dieci ballate. Il modello di elaborazione del linguaggio naturale imparerà gli stili di ciascuno dei personaggi utilizzando le loro dieci novelle. Se il modello può quindi prevedere l'autore delle ballate con una precisione superiore al 10%, allora le ballate rientrano nella caratterizzazione stilistica della brigata. Poiché le ballate, una forma di poesia, hanno lo scopo di rivelare la psiche emotiva dell'autore, questa è un'ipotesi logica. Le ballate creano un senso di ordine nel *Decameron* a causa del ruolo integrale che hanno nel dare potere al sovrano di ogni giornata e perché si adattano alle strutture stilistiche di ciascuno dei membri. Panfilo disponeva del set di dati di addestramento più ampio, quindi ci si aspettava che i risultati della previsione delle sue ballate sarebbero stati i più accurati. Tuttavia, Neifile e Filostrato hanno avuto i risultati di previsione più accurati. Ci si aspettava che Neifile avrebbe ottenuto i risultati meno accurati perché il suo training set era di qualche paragrafo in meno rispetto agli altri membri. Invece Lauretta ha avuto i risultati peggiori.

### *Metodologia di Classificazione del Testo*

Ci sono sei passaggi per creare programmi di classificazione di un testo:<sup>40</sup>

1. Creare un set di dati etichettato. Etichettato significa che i dati devono essere già classificati.

---

<sup>40</sup> Vajjala, "4. Text Classification - Practical Natural Language Processing [Book]."

2. Suddividere il set di dati in dati di addestramento e dati di test e stabilire una metrica delle prestazioni. I dati di addestramento vengono utilizzati per aiutare il programma a imparare come devono essere classificati i diversi testi. Questo è il motivo per cui il set di dati dovrebbe essere preclassificato. I dati del test vengono utilizzati per testare l'andamento dell'allenamento. Ancora una volta, affinché funzioni, è importante che il set di dati sia stato preclassificato. Una metrica delle prestazioni fa riferimento a un punteggio che puoi utilizzare per valutare quanto bene il programma ha appreso i dati di addestramento.
3. Trasformare il testo in un vettore di caratteristiche. Un vettore di caratteristiche è il testo ma in un formato di elenco che il computer può comprendere.
4. Addestrare un modello utilizzando i dati di addestramento.
5. Utilizzando la metrica delle prestazioni, testare il modello con i dati di test.
6. Distribuire il modello su set di dati al di fuori dei dati di addestramento e test e verificare le prestazioni.

È buona regola eseguire i passaggi 3-5 molte volte, in modo da scegliere la migliore variazione del vettore di caratteristiche. Forse il testo dovrebbe essere suddiviso di più? Forse non ci sono abbastanza dati? Esistono molti modi per personalizzare un vettore di caratteristiche. E necessario tenere presente che questi passaggi possono essere utilizzati solo in un modello di apprendimento supervisionato. Il modello di apprendimento supervisionato è quando le classi sono predeterminate. Nel modello di apprendimento non supervisionato, la classe dei dati di input è sconosciuta.<sup>41</sup>

---

<sup>41</sup> Vajjala.

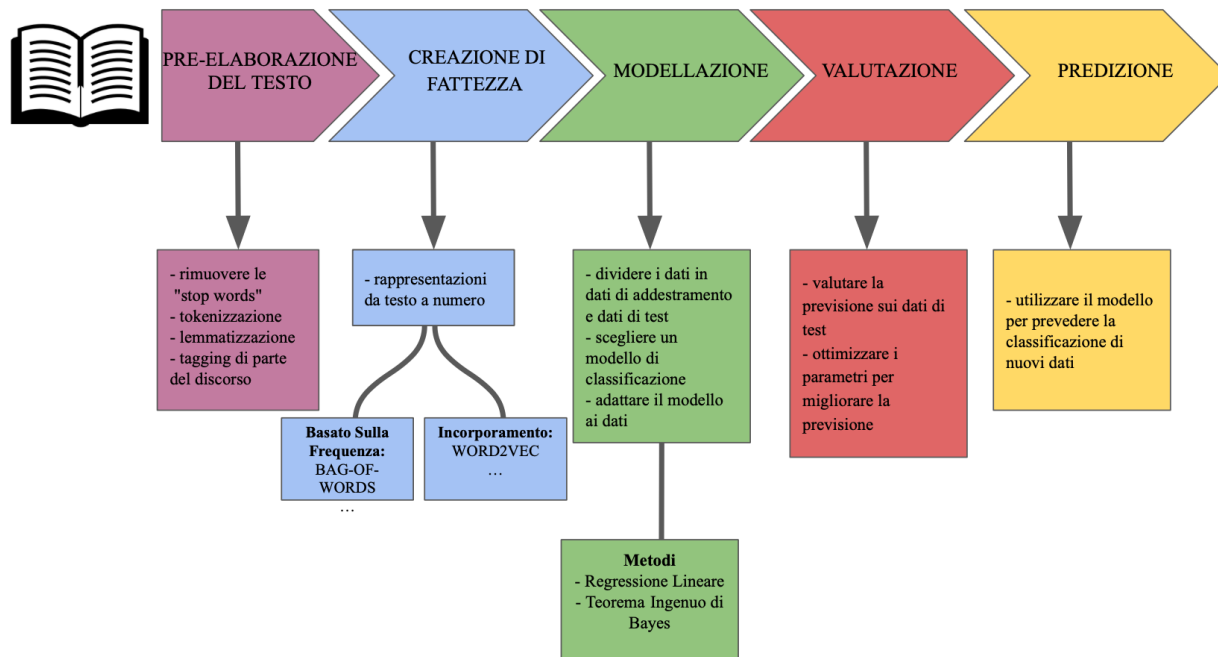


Figura 6. Pipeline di linguaggio naturale di base

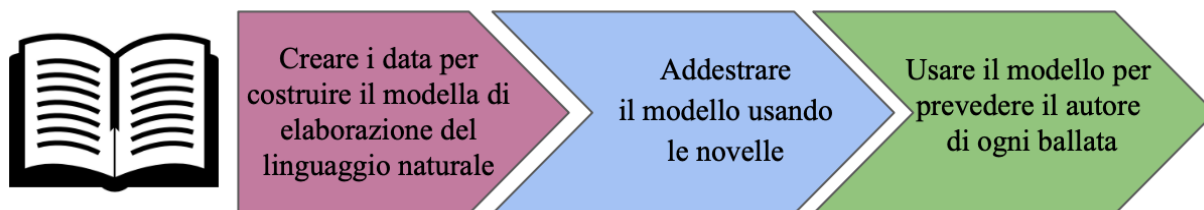


Figura 7. Pipeline ELN generale per la previsione dell'autore delle dieci ballate.

### Usando il ELN per mostrare l'ordine nel Decameron

La congruenza degli stili di scrittura tra le storie di ciascun personaggio e il rispetto della ballata stabilizzano l'ordine della struttura narrativa del *Decameron*? Oppure le canzoni deviano dallo stile del testo e destabilizzano il senso dell'ordine? A queste domande verrà data risposta attraverso l'ELN. Seguendo la pipeline di classificazione del testo impostata all'inizio di questo capitolo, il problema verrà impostato utilizzando un sistema di classificazione multiclasse. Le classi sono le seguenti:

- 1) Pampinea

- 2) Filomena
- 3) Neifile
- 4) Filostrato
- 5) Fiammetta
- 6) Elissa
- 7) Dioneo
- 8) Lauretta
- 9) Panfilo
- 10) Emilia

La classificazione multiclasse sarà impostata utilizzando la pipeline stabilita sopra. Questi sono i parametri specifici:

1. Le 100 novelle saranno utilizzate come set di dati per il test.
2. Si puliscono i dati del test rimuovendo tutte le stop word e sostituendo la punteggiatura con un'etichetta appropriata.
3. Il 75% dei paragrafi da 100 novelle verrà utilizzato come dati di addestramento. Le novelle saranno pre-classificate utilizzando i rispettivi autori.
4. Come dati del test verrà utilizzato un venticinque per cento randomizzato del testo.
5. La metrica delle prestazioni sarà determinata in base alla capacità o meno del modello di classificare correttamente i dati del test.
6. I dati verranno trasformati in vettori di caratteristiche.
7. Il modello verrà quindi addestrato utilizzando i dati di addestramento. Questo passaggio sarà il più lungo.

8. Il modello verrà quindi testato utilizzando i dati del test. Se il modello è in grado di classificare correttamente i dati del test, allora il modello può provare a classificare la ballata.
9. I passaggi da sei a otto verranno ripetuti fino al raggiungimento della corretta classificazione dei dati del test.

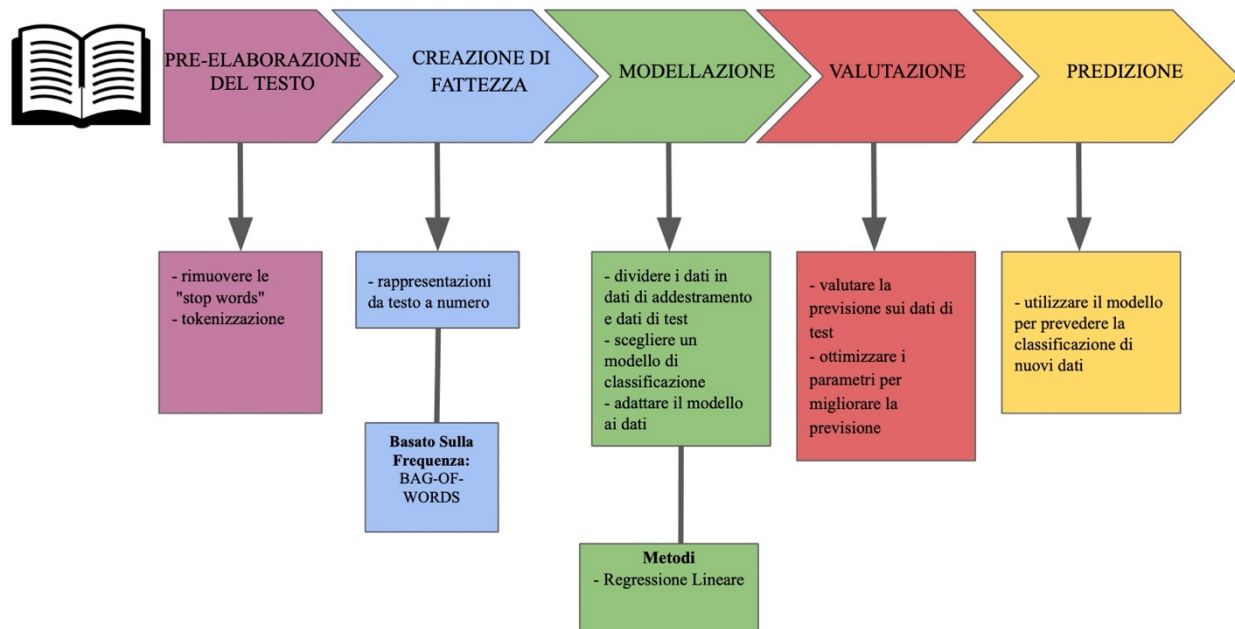


Figura 8. La pipeline ELN che è stata effettivamente utilizzata per la previsione della paternità delle ballate

*Scelta di rimuovere la lemmatizzazione e il tagging di parte del discorso dal processo di pre-elaborazione del testo*

Ho rimosso la lemmatizzazione e la parte del discorso dal tagging della pipeline ELN nel

*Decameron* perché la libreria python che ho usato non supportava queste funzionalità

nell'italiano medievale. La lemmatizzazione è una tecnica di elaborazione del linguaggio naturale che comporta la riduzione di una parola alla sua forma base o dizionario, chiamata lemma.

L'obiettivo della lemmatizzazione è normalizzare le parole in modo che le diverse forme della stessa parola siano trattate come lo stesso token. Ad esempio, in italiano, il verbo "parlare" può assumere varie forme a seconda del tempo, dell'umore e della persona, come "parlo" (io parlo),

"parli" (tu parli), "parla" (lui/lei parla), "parliamo" (parliamo), e così via. Lemmatizzando queste forme, esse vengono ridotte alla loro forma base "parlare". La lemmatizzazione è utile in molte attività di elaborazione del linguaggio naturale, come la classificazione del testo, il recupero delle informazioni e la traduzione automatica. Normalizzare le parole alla loro forma base riduce la complessità del vocabolario e rende più facile identificare il significato e le relazioni tra le parole.

Il tagging di parte del discorso (POS) è un'attività di elaborazione del linguaggio naturale che comporta l'assegnazione di ogni parola in un testo a una categoria grammaticale, come sostantivo, verbo, aggettivo, avverbio o preposizione. L'obiettivo del POS tagging è identificare il ruolo che ogni parola gioca in una frase e facilitare un'ulteriore analisi del testo. In italiano, come in molte altre lingue, le parole possono avere molteplici possibili tag POS a seconda del contesto in cui compaiono. Ad esempio, la parola "sono" può essere una forma del verbo "essere" (essere) alla prima persona singolare presente ("io sono"), oppure può essere una ausiliare che significa "io sono" ("sono arrivato" significa "sono arrivato"). Per eseguire il POS tagging per l'italiano, è possibile utilizzare varie tecniche e strumenti, come sistemi basati su regole, modelli statistici e algoritmi di apprendimento automatico. Questi metodi si basano su una combinazione di conoscenze linguistiche, come regole grammaticali e lessici, e modelli statistici osservati in corpora testuali di grandi dimensioni. Mentre la lemmatizzazione e l'etichettatura di parte del discorso sono utili, il loro mancato utilizzo non crea un modello di linguaggio naturale scadente. Un modello può ancora produrre risultati accurati ma sarà più efficace nel riconoscere le classi con stili anomali (ad es. molto tristi o molto felici).



### *Scelta di Word2Vec e Regressione Lineare*

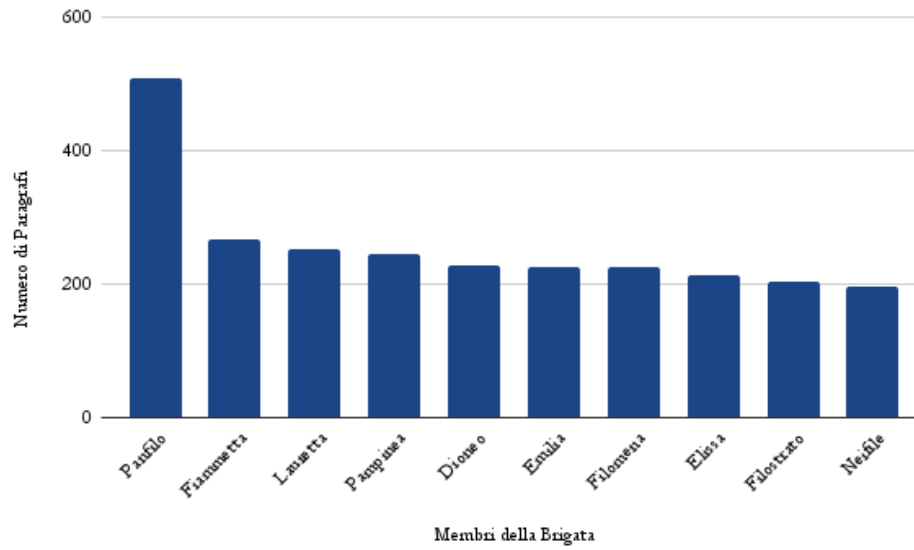
La scelta tra i modelli Bag-of-Words (BoW) e Word2Vec dipende dall'attività specifica e dalla natura dei dati. Entrambi i modelli hanno i loro punti di forza e di debolezza e il modello più adatto dipenderà dai requisiti dello specifico compito di PNL. Bag of Words è un modello più semplice e diretto rispetto a Word2Vec. È utile per attività che richiedono una semplice rappresentazione del testo e non sono sensibili al contesto e all'ordine delle parole. Ad esempio, BoW viene spesso utilizzato per l'analisi del sentimento e le attività di classificazione del testo, in cui la presenza o l'assenza di parole specifiche è più importante dell'ordine delle parole. D'altra parte, Word2Vec è un modello più sofisticato che tiene conto del contesto e dell'ordine delle parole nel testo. È utile per le attività che richiedono una comprensione più sfumata del significato del testo, come la modellazione linguistica, la traduzione automatica e il recupero delle informazioni. Word2Vec è un strumento migliore per catturare le relazioni tra le parole e le sfumature dei loro significati. Riassumendo, se l'attività richiede una semplice rappresentazione del testo e l'ordine delle parole non è critico, Bag of Words può essere una scelta adatta. Tuttavia, se l'attività richiede una comprensione più sfumata del testo e delle relazioni tra le parole, allora Word2Vec potrebbe essere una scelta migliore. In definitiva, la scelta tra questi modelli dipenderà dai requisiti specifici dell'attività di ELN a disposizione.

Ho scelto Word2Vec perché è un metodo basato sull'incorporamento che cattura più sfumature rispetto a Bag-of-Words. Word2Vec è un metodo basato sull'incorporamento perché mappa le parole da uno spazio ad alta dimensione, come lo spazio di tutte le parole possibili in una lingua, a uno spazio a dimensione inferiore, tipicamente poche centinaia di dimensioni, dove ogni parola è rappresentata come vettore denso di numeri reali. Questa rappresentazione vettoriale è nota

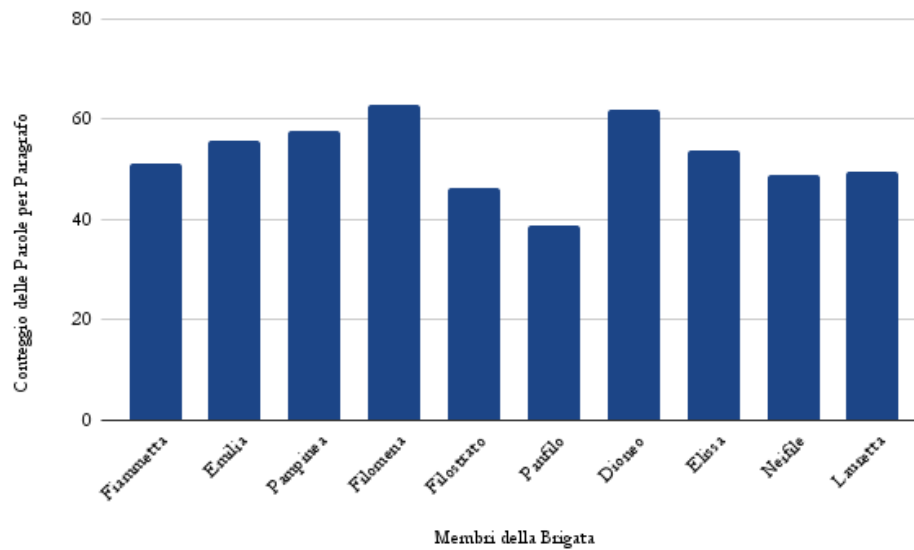
come incorporamento di parole. L'idea alla base degli incorporamenti di parole è catturare il significato delle parole rappresentandole in un modo che catturi le relazioni tra di loro.

Posizionando parole simili più vicine l'una all'altra nello spazio di incorporamento e parole dissimili più lontane, gli incorporamenti di parole forniscono una rappresentazione più efficace delle parole per molte attività di elaborazione del linguaggio naturale. Word2Vec apprende gli incorporamenti di parole addestrando una rete neurale su un ampio corpus di testo. La rete è addestrata a prevedere il contesto di una data parola o, in alternativa, a prevedere la parola dato il suo contesto. Questo processo di addestramento della rete neurale si traduce in incorporamenti di parole che catturano le proprietà distributive delle parole nel corpus, consentendo alla rete di catturare relazioni semantiche e sintattiche tra le parole. Gli incorporamenti di parole appresi da Word2Vec sono densi e poco dimensionali, il che li rende computazionalmente efficienti da utilizzare nelle attività ELN a valle, come la classificazione del testo, il recupero delle informazioni e la traduzione automatica. Inoltre, gli incorporamenti possono essere utilizzati per inizializzare i pesi delle reti neurali profonde, il che ha dimostrato di migliorare le prestazioni di queste reti in molte attività di ELN. In sintesi, Word2Vec è un metodo basato sull'incorporamento perché apprende una rappresentazione vettoriale a bassa dimensione delle parole che ne cattura il significato e le relazioni in un dato corpus di testo.

*Metriche dei Dati di Addestramento*



*Figura 9. Numero di Paragrafi per ogni membro della brigata.*



*Figura 10. Numero di parole per ogni membro della brigata.*

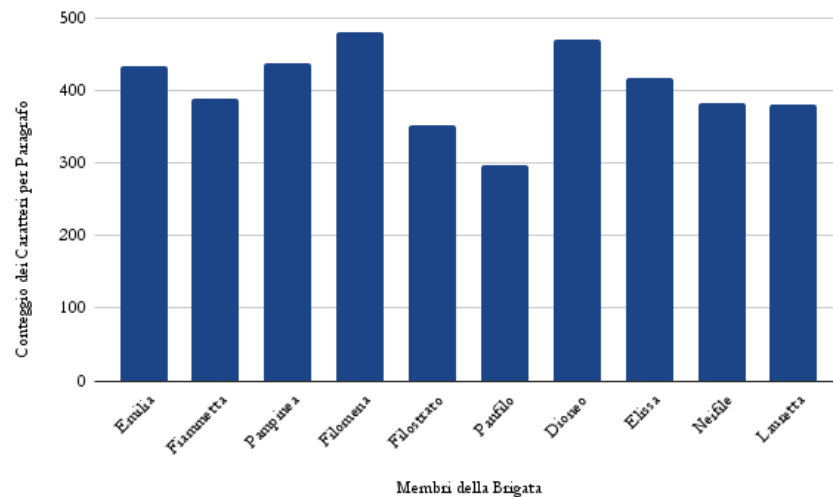


Figura 11. Numero di caratteri per ogni membro della brigata.

L'accuratezza dei dati di addestramento si basa su due fattori principali: (1) la qualità dei dati e (2) le somiglianze stilistiche tra le novelle di un membro e la loro ballata. Avere una comprensione dei dati che vengono creati fornirà un quadro migliore del primo fattore. I dati dell'addestramento sono stati creati suddividendo ciascuna delle novelle del membro della brigata per paragrafo. I dati sono stati suddivisi per paragrafo perché le ballate hanno all'incirca le dimensioni di un paragrafo. Avere dati di addestramento di dimensioni simili aiuta con l'accuratezza della previsione del modello.<sup>42</sup> Panfilo ha il maggior numero di paragrafi nel suo dataset (Fig. 4), mentre Neifile ha il minor numero di paragrafi nel suo set di dati (Fig. 4). Anche se Panfilo ha il maggior numero di paragrafi, ha il minor numero di parole e caratteri all'interno di quei paragrafi (Fig. 5-6). Senza considerare Panfilo, gli altri membri della brigata hanno un numero medio di paragrafi di circa 228 (Fig. 4). Ciò significa che nel set di dati di addestramento, ogni membro della brigata ha circa 228 voci di dati. Poiché Neifile ha il numero

<sup>42</sup> Vajjala.

di paragrafi più basso, i suoi dati sono più suscettibili all'overtraining. Pertanto, è probabile che il modello abbia difficoltà a prevedere l'autore della sua ballata.

*Predizione di Autore delle Ballate*

|            | Precisione | Recall | F1-Score | Support |
|------------|------------|--------|----------|---------|
| Fiammetta  | 0.59       | 0.68   | 0.63     | 56      |
| Emilia     | 0.63       | 0.49   | 0.55     | 35      |
| Pampinea   | 0.54       | 0.52   | 0.53     | 48      |
| Filomena   | 0.60       | 0.62   | 0.61     | 45      |
| Filostrato | 0.81       | 0.49   | 0.61     | 43      |
| Panfilo    | 0.57       | 0.92   | 0.70     | 105     |
| Dioneo     | 0.67       | 0.65   | 0.66     | 46      |
| Elissa     | 0.70       | 0.34   | 0.46     | 47      |
| Neifile    | 0.82       | 0.44   | 0.57     | 41      |
| Lauretta   | 0.52       | 0.47   | 0.49     | 47      |

43

Tabella 1. Risultati della previsione del modello di regressione lineare dell'autore della ballata.

La colonna di dati più preziosa è la colonna di precisione. Il modello è stato in grado di prevedere correttamente l'autore di ciascuna delle ballate oltre il 50%, il che suggerisce che ciascuno dei membri della brigata mantiene il proprio stilome tra la novella e la ballata (Tabella 1). Un valore di precisione superiore al 10% suggerisce che il modello non ha indovinato a caso

---

<sup>43</sup>Precisione, F1, richiamo e supporto sono metriche comunemente utilizzate nel campo dell'apprendimento automatico per valutare le prestazioni di un modello di classificazione. Queste metriche sono particolarmente utili quando si ha a che fare con set di dati sbilanciati, in cui una classe può avere molte più istanze di un'altra. Queste metriche possono essere utilizzate per valutare le prestazioni di un modello tassonomico

Accuratezza: l'accuratezza è il rapporto tra il numero di istanze previste correttamente e il numero totale di istanze nel set di dati.

F1: F1 è la media armonica di precisione e richiamo. La precisione è il rapporto tra i veri positivi e il numero totale di previsioni positive, mentre il richiamo è il rapporto tra i veri positivi e il numero totale di istanze positive effettive nel set di dati. F1 è una metrica più bilanciata dell'accuratezza quando si tratta di set di dati sbilanciati, poiché tiene conto sia della precisione che del richiamo.

Richiamo: il richiamo, noto anche come sensibilità, è il rapporto tra i veri positivi e il numero totale di istanze positive effettive nel set di dati.

Supporto: il supporto è il numero di istanze nel set di dati che appartengono a una particolare categoria. (Vajjala.)

e che esiste uno stilome unico per ogni personaggio. Questa scoperta contribuisce al senso dell'ordine nel *Decameron* e afferma che le ballate si inseriscono nella struttura narrativa del testo. Le ballate giocano già un ruolo importante nella struttura narrativa del *Decameron*. Assegnare un cantastorie è il primo compito quotidiano del sovrano. Questa scoperta, insieme alla posizione delle ballate nella struttura narrativa del *Decameron*, consolida il ruolo fondamentale che le ballate svolgano nel libro. I risultati della previsione dell'autore della ballata di Panfilo, Neifile, Filostrato e Lauretta hanno prodotto i risultati più inaspettati e richiedono ulteriori analisi.

### *Analisi di Panfilo*

I risultati di precisione ELN di Panfilo hanno deviato dai valori attesi a causa delle dimensioni dei suoi dati di allenamento. Panfilo ha un recall significativamente più alto rispetto agli altri membri della brigata (Tabella 1). Ciò significa che di tutte le previsioni corrette fatte dal modello, il 92% delle previsioni corrette erano per la ballata di Panfilo (Tabella 1). Questo perché Panfilo aveva molte più voci di dati nel set di dati di addestramento. Tuttavia, il suo valore di precisione (57%) è il secondo più basso di tutti i membri della brigata. Ciò potrebbe suggerire che Panfilo non abbia uno stilome molto consistente attraverso la sua novella e ballata rispetto agli altri membri. Tuttavia, l'analisi della sua ballata nel capitolo precedente rivela che l'ambiguità della poesia di Panfilo era un allontanamento dalla sua scrittura precedente e confondeva persino gli altri membri della brigata.

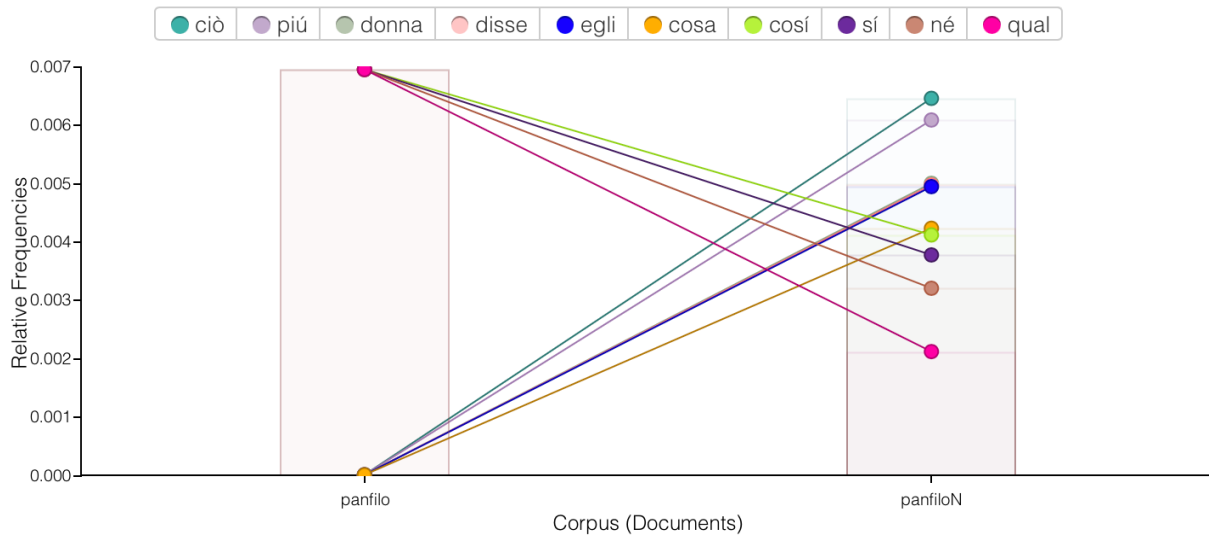


Figura 12. Dieci parole con la frequenza più alta nelle Novelle di Panfilo (a destra) Mappate con la sua ballata (a sinistra)

Un'analisi delle dieci parole più frequenti nelle novelle associate alla ballata di Panfilo può fornire un'idea della mancanza di precisione. ‘Ciò’, ‘più’ e ‘donna’, le parole più frequenti nella novella di Panfilo non compaiono una volta nella sua ballata (Fig. 7). Questo squilibrio nelle parole ad alta frequenza tra i due corpus potrebbe iniziare a spiegare il tasso di precisione inferiore di Panfilo rispetto ad altri membri della brigata. Una lettura attenta delle sue novelle in relazione alla sua ballata potrebbe fornire maggiori informazioni su questa scoperta.

### *Analisi di Neifile*

I risultati della previsione ELN di regressione lineare per la ballata di Neifile hanno deviato dalle aspettative. Poiché Neifile aveva il numero più basso di voci nei dati di addestramento, il modello era suscettibile di sovrallenamento dei suoi dati (Fig. 4). L'overtraining si verifica quando non ci sono abbastanza voci di dati per una determinata classe e il modello memorizza le informazioni invece di apprendere. Ciò può impedire al modello di gestire informazioni o parole che non sono presenti nei dati di addestramento. Inaspettatamente, il modello ELN è stato in



grado di prevedere correttamente che Neifile ha cantato la sua ballata per l'82% delle prove eseguite (Tabella 1). Poiché Neifile aveva il numero più basso di voci nel set di dati di addestramento, il suo valore di richiamo è il più basso al 44% (Tabella 1). Questo valore non riflette una minore continuità tra la sua ballata e le sue novelle, ma piuttosto la minore dimensione del suo sottoinsieme di dati di addestramento rispetto agli altri membri della brigata. Per capire perché la Precisione di Neifile fosse così alta, è necessario confrontare i dati di frequenza delle parole tra le novelle di Neifile e la sua ballata.

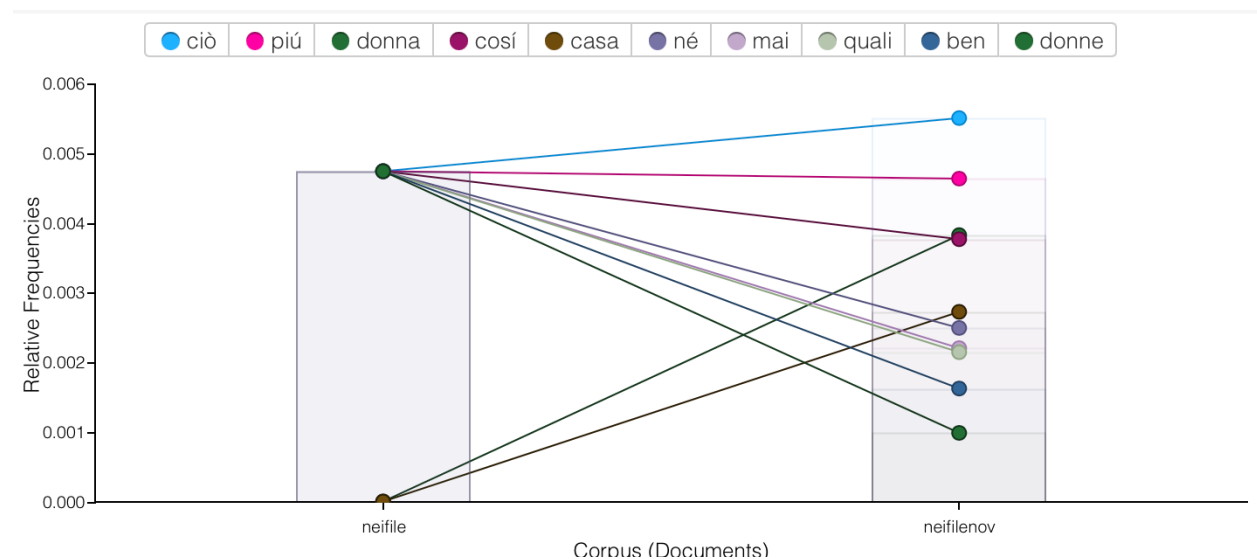


Figura 13. Dieci parole con la frequenza più alta nelle Novelle di Neifile (a destra) Mappate con la sua ballata (a sinistra)

Un'analisi delle dieci parole più frequenti può dare un'idea del valore di precisione di Neifile. Solo due delle parole più frequenti nella novella di Neifile non sono presenti nella sua ballata. Le due parole che non sono presenti sono 'casa' e 'donne' (Fig. 8). È interessante che le otto parole presenti anche nella ballata di Neifile siano presenti con frequenza simile nella ballata. Sebbene sia difficile trarre conclusioni dirette sulla somiglianza nell'argomento senza fare un'analisi più approfondita delle novelle di Neifile, l'accuratezza del modello nel prevedere l'autore della

ballata di Neifile suggerisce continuità nella voce di Neifile in entrambi i corpus. Le sue ballate e novelle rivelano l'ordine nel suo corpus narrativo.

### *Analisi di Filostrato*

Filostrato ha avuto il secondo tasso di precisione più alto (81%) dei membri della brigata, rivelando una continuità nella sua struttura narrativa attraverso la sua novella e ballata (Tabella 1). Questa è una scoperta interessante perché le informazioni sul suo set di dati di addestramento hanno dato motivo di credere che il modello di apprendimento avrebbe avuto più difficoltà ad apprendere il suo stilome. Il conteggio delle parole dei dati inseriti da Filostrato era inferiore a quello degli altri membri della brigata. Questa realtà si riflette nel suo tasso di richiamo inferiore alla media del 49%. Il basso tasso di richiamo è un riflesso di paragrafi più brevi nella novella di Filostrato. La qualità dei suoi dati di addestramento non è stata danneggiata perché aveva ancora una quantità di dati immessi simile a quella degli altri membri della brigata. Uno sguardo ai dati di frequenza per la ballata e la novella di Filostrato può fornire informazioni sul suo tasso di alta precisione.

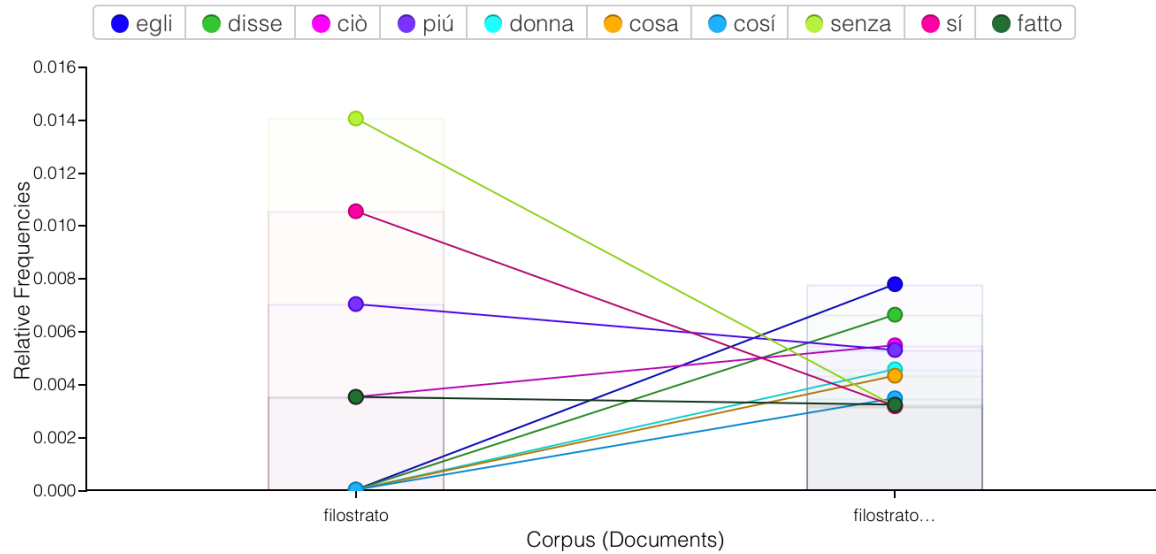


Figura 14. Dieci parole con la frequenza più alta nelle Novelle di Filostrato (a destra) Mappate con la sua ballata (a sinistra).

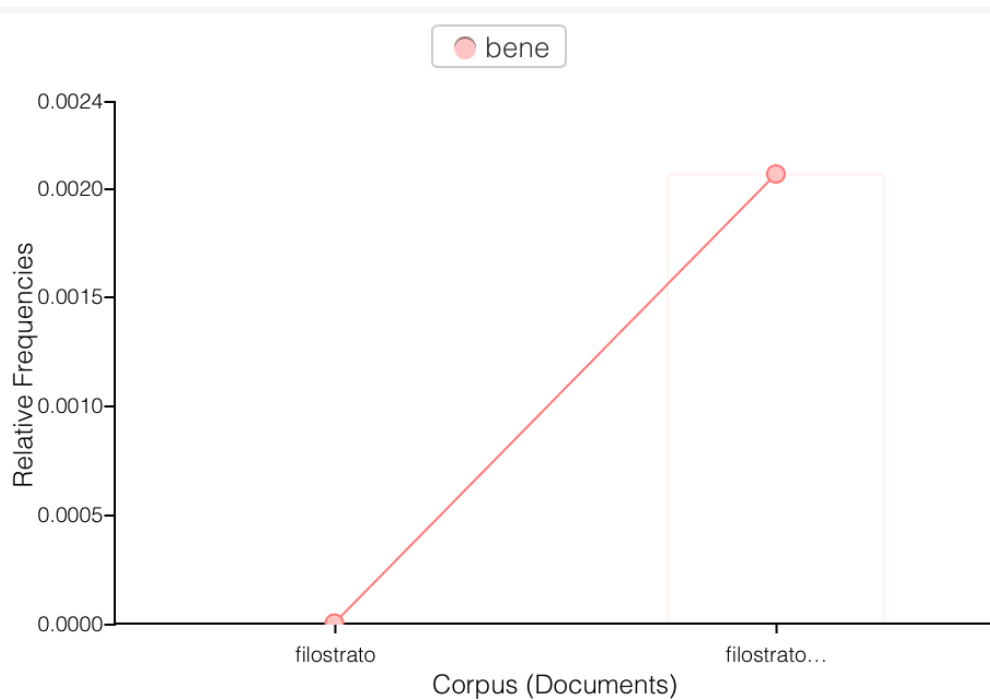


Figura 15. Frequenza di “bene” nella novella di Filostrato mappata alla sua ballata. La parola “bene” appare molte volte nella sua, ma mai nella sua ballata.

La mappatura delle dieci parole con la frequenza più alta nella novella di Filostrato rispetto alla sua ballata rivela un alto senso dell'ordine tra i suoi due corpera. Cinque delle parole con la frequenza più alta nella sua novella sono presenti nella ballata (Fig. 9). "Senza" è effettivamente

presente nella sua ballata con circa il doppio della frequenza che ha nella novella. L'analisi precedente della ballata di Filostrato ha caratterizzato la sua canzone con una profonda tristezza. L'alta prevalenza della parola "senza" in entrambe le opere parla dell'ordine della sua struttura narrativa e dimostra come la psiche emotiva permei la sua novella e non solo la sua ballata. Dopo aver trovato questo, ho deciso di guardare la frequenza di una parola che rappresenta l'opposto della tristezza: "bene". Sebbene "bene" fosse piuttosto diffuso nella sua novella, non è mai apparso nella sua ballata (Fig. 10). Ciò riflette il potere della ballata come mezzo che consente la più vera espressione emotiva. Sebbene l'alta frequenza di "senza" mostri continuità nella voce, l'assenza della parola "bene" nella sua ballata rivela come Filostrato fosse in grado di esprimere la vera portata della sua angoscia nella sua ballata.

#### *Analisi di Lauretta*

I risultati dell'ELN rivelano che Lauretta ha il peggior tasso di precisione (52%) della brigata (Tabella 1). Ciò significa che la sua ballata non si adatta bene allo stilome creato nelle sue novelle rispetto agli altri membri della brigata. Tuttavia, è importante notare che poiché il valore di precisione è superiore al 10%, il modello ELN è ancora in grado di riconoscere uno stilome distintivo. Osservando le metriche del suo set di dati di addestramento, non ci sono valori degni di nota che indichino la sua bassa percentuale di precisione (Fig. 4-6). Una visione delle parole con la frequenza più alta nelle sue novelle mappate sulla sua ballata potrebbe fornire un'idea del suo basso valore di precisione.

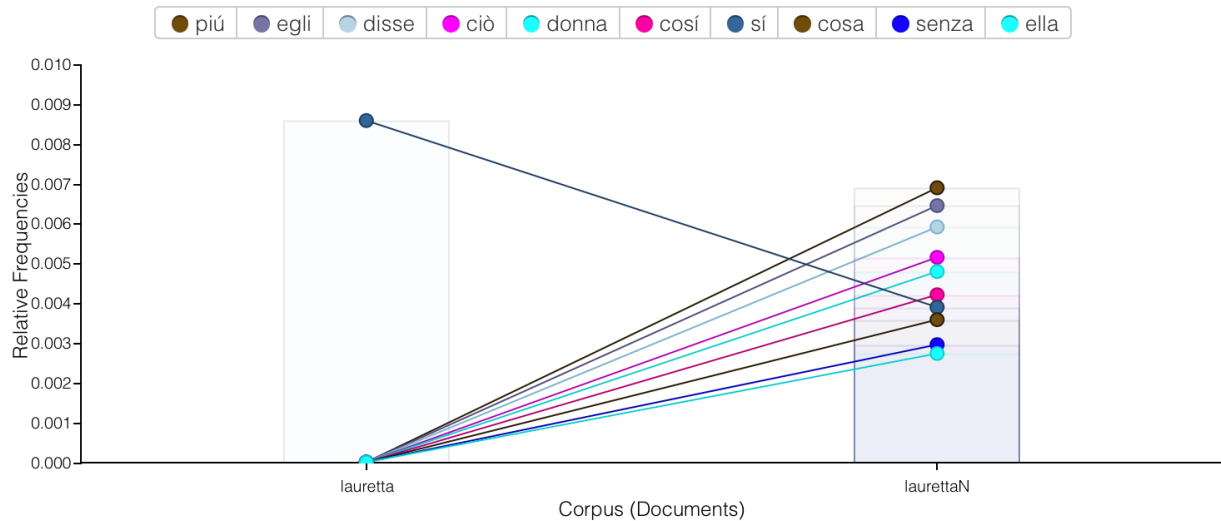


Figura 16. Dieci parole più frequenti nelle Novelle di Lauretta (a destra) Mappate con la sua ballata (a sinistra)

Solo una delle parole più frequenti nella novella di Lauretta è presente nella sua ballata. Sebbene ciò non indichi definitivamente una dissonanza soggettiva tra i due corpora, suggerisce che il suo stilome nel romanzo sia creato da scelte stilistiche più sottili, come l'uso della punteggiatura e il fraseggio non contenuto. Uno studio approfondito delle sue novelle rispetto alla sua ballata è necessario per trarre conclusioni importanti. Tuttavia, il valore di precisione del 52% parla dello stilome distintivo di Lauretta. Non è così distinguibile come le voci degli altri membri della brigata.

### Riassunto

L'uso della elaborazione della lingua naturale dimostra che le ballate di tutti i membri della brigata si inseriscono nelle strutture narrative di ciascuna delle loro novelle. Per ottenere questi risultati è stato utilizzato un modello di apprendimento della regressione lineare. Nonostante le dimensioni superiori alla media del sottoinsieme dei dati di addestramento di Panfilo, il modello non è stato in grado di prevedere al meglio l'autore della sua ballata. Invece Neifile, il membro

della brigata con il sottoinsieme di dati di addestramento più piccolo, aveva la maggiore precisione nel prevedere l'autore della ballata. Questo fatto si rifletteva nella continuità delle parole ad alta frequenza tra la novella e la ballata. Filostrato aveva la seconda ballata più prevedibile, e Lauretta aveva la ballata meno prevedibile. A livello narrativo, questi risultati riflettono l'ordine che queste ballate creano tra le novelle della brigata e la loro psiche emotiva complessiva. Come forma di poesia, le ballate esistono per permettere al lettore di scrutare nelle menti della brigata. L'accuratezza della previsione della ballata inferiore alla media non riflette solo la dissonanza tra novella e ballata. Rivela anche come gli esseri umani custodiscono le emozioni quando interagiscono con gli altri e come la poesia abbatte le barriere emotive e consente alla psiche emotiva di venire alle luce. Piuttosto che fungere da allontanamento dalla struttura narrativa regolare creata dalle novelle, le ballate ristabiliscono l'ordine nel *Decameron*, anche letteralmente nel programma del giornata del *Decameron*. Le ballate segnano la fine della giornata e l'inizio di un nuovo regno e ordine in un senso più poetico. Ad ogni membro della brigata viene data l'opportunità di ordinare le proprie emozioni attraverso il canto.

## Conclusione

### *Riassunto dei Risultati*

Questo progetto di discipline umanistiche digitali ha sottolineato come i computer, un moderno progresso tecnologico, siano compagni naturali e necessari per chiunque cerchi di comprendere il

*Decameron*, un importante testo medievale definitivo scritto nel Trecento. La stilometria ha facilitato la comprensione della complessa psiche emotiva della brigata. L'elaborazione della linguaggio naturale ha potuto mostrare come le ballate si allineino con i stilomi creati nelle novelle di ogni membro della brigata e ristabiliscano l'ordine alla fine di ogni giornata.

I test di stilometria hanno rivelato come le ballate, una forma musicale di poesia, offrano ai lettori uno sguardo intimo nella complessa psiche emotiva della brigata. In particolare, i test stilometrici hanno reso chiaro il tono cupo della ballata di Panfilo, dimostrato come le ballate presentino le innovative rappresentazioni di genere di Boccaccio e confermato le classificazioni predeterminate delle ballate dei membri della brigata per tono allegro e triste. Sulla base degli stilomi calcolati, i membri della brigata sono stati suddivisi in due gruppi distinti: Gruppo 1: Panfilo, Fiammetta, Filomena e Gruppo 2: Emilia, Pampinea, Neifile, Dioneo, Lauretta, Elissa e Filostrato. Panfilo e Filostrato, i due personaggi maschili più vicini per età, cantano ballate che hanno le voci stilistiche più distinte. Che se ne fosse reso conto o meno, Boccaccio ha mostrato ai suoi lettori del Trecento quanto possano essere accidentati i paesaggi emotivi e come non rispettino le normali linee di genere. Invece, sono guidati da esperienze umane idiosincratice con l'amore. La ballata di Panfilo si allinea maggiormente con due donne tristi: Filomena, un'amante nostalgica, e Fiammetta, una fidanzata gelosa. Nel Gruppo 2 il test stilometrico ha potuto caratterizzare correttamente come allegre le ballate di Emilia, Pampinea, Neifile e tristi

quelle di Dioneo, Lauretta, Elissa e Filostrato. L'analisi letteraria ha rivelato che il confine tra felice e triste è stato superato dalle belle immagini della natura nelle ballate di Neifile e Dioneo.

L'elaborazione del linguaggio naturale ha rivelato che le ballate non sono solo dolci canzoni che la brigata canta dopo cena. Un modello di apprendimento è stato in grado di identificare la voce unica di ogni membro della brigata e utilizzare queste informazioni per prevedere correttamente l'autore di ciascuna delle ballate. Un metodo di vettorizzazione word2vec è stato usato invece del metodo Bag-of-Words nel tentativo di aggiungere più sfumature e conservare quante più informazioni possibili dal testo originale. È stato utilizzato un modello di apprendimento di regressione lineare rispetto al Teorema di Bayes perché può accogliere parole e frasi non presenti nel set di dati di addestramento.

### *Impatto dello Studio sul Decameron*

I risultati presentati in questo studio affermano Boccaccio come uno straordinario scrittore innovativo del Trecento e confermano il valore delle ballate come soggetti di ricerca. I test stilometrici e l'elaborazione del linguaggio naturale incentrato sulle ballate hanno rivelato la padronanza da parte di Boccaccio di ciascuna delle voci uniche dei membri della brigata e la sua profonda capacità di rappresentare la condizione umana in una bella prosa. In particolare, l'accorpamento stilometrico delle ballate ha potuto mettere in luce il carattere ambivalente della ballata di Panfilo. Neanche gli altri membri della brigata sapevano come reagire alla sua ballata. Sebbene fosse piena di rappresentazioni positive dell'amore, la sua incapacità di trasmettere adeguatamente le sue emozioni e l'identità del suo amante lo rattrista. Poiché i test stilometrici hanno raggruppato la ballata di Panfilo con Filomena e Fiammetta, è possibile una nuova lettura della ballata di Panfilo. Le ballate di Filomena e Fiammetta raffigurano le loro malinconiche



emozioni sull'amore. La ballata di Filomena è piena di nostalgia disperata per l'amore passato, ricorda il suo amante e desidera il loro ricongiungimento. Fiammetta esprime rabbia gelosa: è innamorata, ma il fuoco della sua gelosia nei confronti dell'amato la consuma. La sua triste insicurezza minaccia la stabilità della sua relazione. Dal momento che Panfilo è raggruppato con loro diventa chiaro che la sua incapacità di trasmettere le sue emozioni maschera il dubbio che nutre per le prospettive del suo amore appassionato. Nella sua generosità, si preoccupa di cosa potrebbe diventare in sua assenza da lei. Senza stilometria, questa lettura della sua ballata da parte sua non avrebbe tanto merito. Poter leggere la ballata in relazione ad altre ballate rende possibili nuovi punti di vista. Elaborazione del linguaggio naturale ha dato una chiara visione della capacità di Boccaccio di mantenere le voci dei membri della brigata nelle loro novelle e ballate. Questo parla del grande talento di Boccaccio ma anche della tradizione orale di molte delle storie raccontate nel *Decameron*.<sup>44</sup> Boccaccio ha svolto il difficile compito di scrivere le novelle e le ballate, ma è possibile che le voci di alcuni degli oratori originali siano presenti nelle voci dei personaggi del *Decameron*. La capacità dell'ELN di prevedere la ballata supporta questa speculazione.

### *Miglioramento*

Un progetto di umanistica digitale incentrato sul *Decameron* presenta alcune sfide. Sebbene il *Decameron* sia un testo interessante, è scritto in italiano medievale, che non è supportato dalle librerie Python utilizzate nel modello ELN. Pertanto, la lemmatizzazione e l'etichettatura di parte del discorso non sono state completate. Se le biblioteche potessero essere ampliate per accogliere l'italiano medievale, questi processi potrebbero essere eseguiti producendo risultati più sfumati.

---

<sup>44</sup> Kuhns, *Decameron and the Philosophy of Storytelling: Author as Midwife and Pimp*, 15.

### *Ulteriori studi*

Va notato che gli strumenti computazionali utilizzati in questo progetto sono i migliori per identificare i valori anomali. Gli algoritmi di stilometria e classificazione analizzano i modelli in grandi insiemi di dati per identificare somiglianze e differenze tra loro. Valori anomali nei dati, o punti dati che non rientrano nell'intervallo di valori previsto, possono fornire informazioni preziose e potenzialmente rivelare nuovi modelli o relazioni nei dati. Questo è il motivo per cui molti degli stessi membri della brigata sono stati evidenziati nei reperti di stilometria e elaborazione del linguaggio naturale. Pertanto, tre donne della brigata, Elissa, Emilia e Pampinea, non hanno ricevuto un'attenzione significativa. Sebbene non siano valori anomali, le loro voci sono comunque rappresentative dell'esperienza italiana medievale. Un progetto che si concentra sullo sviluppo delle loro voci nelle loro novelle e brigata sarebbe interessante.

I risultati di questo studio sono stati ottenuti utilizzando l'edizione Branca del *Decameron*. Sarebbe interessante vedere se i risultati sono conservati in diverse edizioni e traduzioni del testo. Poiché i traduttori spesso inseriscono il proprio contesto storico nel loro lavoro, molto probabilmente ci sarebbero delle differenze. Queste discrepanze potrebbero fornire informazioni su come le diverse generazioni hanno interagito con il *Decameron*.

### *Il Mio Cammino Verso il Decameron*

Cinque... quattro... tre... due... e puntualissima la signorina Enright è entrata nel mio corso di storia europea dell'AP. Sudata e spaventata, i miei occhi saettavano sui volti di tutti i miei compagni di classe per qualsiasi segno di angoscia. Come si può essere calmi? Nessun altro aveva sentito le storie di notti di studio insonni, esami falliti e GPA offuscati? Avevo sentito in

prima persona i lamenti della mia sorella maggiore mentre lottava per spiegare ai miei genitori perché aveva fatto così male al suo primo test in questa classe. Il mio cuore iniziò a battere forte mentre la signorina Enright continuava a parlare. Avrei subito la stessa sorte di mia sorella? La mia mano faticava a scarabocchiare tutte le parole che uscivano dalla bocca della mia insegnante. “Petrarca... lettere a Cicerone... infatuazione per il mondo che venne prima... umanesimo... Boccaccio...*Decameron*...” Frasi sconnesse volarono alle mie orecchie. Durante una pausa di cinque minuti, mi sono presa un momento per leggere effettivamente i miei scarabocchi. Mentre leggevo, ero sempre più confusa. Questa classe non doveva essere piena di date e facce senza nome? Perché potevo sentire il mio cuore agitarsi mentre leggevo della passione per l'antichità che avevano questi studiosi? Cosa ha spinto questi studiosi ad amare la poesia e l'arte e le persone che l'hanno ispirata? Avevo bisogno di risposte. Dopo la lezione, sono andato ad fontes, tornando alle fonti. Avendo lavorato per anni nelle biblioteche, ho deciso di iniziare con gli autori in ordine alfabetico. Boccaccio. Ho scansionato la sezione B, Bobrowski, Boca, Boccaccio. Le mie dita scivolarono sul dorso di un libro blu brillante. Vorrei poter dire di aver sentito una scarica di elettricità o una sorta di attrazione magnetica. Semmai, ho starnutito per tutta la polvere sul libro che non era mai stato toccato, il libro che aveva aspettato che lo trovassi o, più precisamente, il libro che mi aveva trovato. Il *Decameron*.

Ho aperto il *Decameron*, incerta su cosa aspettarmi. Tentai di leggere il testo sacro ma senza successo. Furiosa, ho sfogliato pagina dopo pagina cercando di capire, faticando a comprendere la poesia che tenevo tra le mani. Mi chiedevo se avessi dimenticato come si legge. Per uno sfortunato spettatore, ero solo una strana ragazza che borbottava tra sé. Potrebbero anche aver pensato che le mie mani agitate fossero un tentativo di lanciare una sorta di incantesimo. Dopo

più tempo di quanto avrei voluto e dopo alcuni falsi incantesimi, mi sono resa conto che il libro era in italiano. Piuttosto che scoraggiarmi, mi convinsi che l'italiano fosse solo spagnolo con un po' di basilico in più. Una volta tornata a casa, ho usato la mia conoscenza dello spagnolo per cercare di decifrare il testo. Passarono solo quindici minuti prima che mi arrendessi e decidessi di ascoltare un audiolibro che avevo trovato online. Mi chiedevo cosa avrebbe pensato Boccaccio. Avrebbe mai potuto prevedere che una ragazza nera di sedici anni del Massachusetts avrebbe premuto la testa contro un rettangolo di metallo per ascoltare il suo libro? Ma mentre ascoltavo le prime pagine, ho sentito un senso di liberazione pervadermi. Erano finite le mie aride e noiose percezioni della vita medievale. La scrittura lirica di Boccaccio mi ha trasportato nella Firenze medievale e l'ho adorato. Boccaccio mi ha fatto ridere dalla comodità del mio letto, in macchina, e fino al liceo. Le suore scandalose e gli amanti infedeli del *Decameron* danzarono fuori dalla pagina, mi fecero qualche capriola nel cervello e mi corsero dritti nel cuore. Le loro lotte e i loro desideri mi hanno ricordato il dramma che ho sentito dagli amici in mensa anche se la brigata aveva respirato oltre 700 anni prima che io fossi nemmeno immaginato. In un tempo e un luogo più antichi di qualsiasi cosa io abbia mai conosciuto. Ma mentre approfondivo il *Decameron*, mi sono reso conto che c'era molto di più che semplici storie divertenti. Boccaccio non era solo un narratore, ma un vero umanista, qualcuno che credeva nel valore innato dell'essere umano e nelle sue capacità. La sua passione per l'antichità e i classici non era solo un semplice fascino per il passato, ma la convinzione che esistessero verità e lezioni senza tempo da apprendere dalle grandi menti del passato. Mi stava conducendo verso la verità su ciò che rende umani gli esseri umani.

Leggendo il *Decameron*, ho potuto sentire l'amore di Boccaccio per l'umanità e la sua fede nel potere della narrazione di connettere le persone attraverso il tempo e lo spazio. Ho capito perché vedeva la letteratura come un modo per elevare e ispirare le persone, per aiutarle a trovare significato e scopo nelle loro vite. Cominciai a capire perché aveva dedicato la sua vita alla poesia, perché aveva scelto di arrendersi al linguaggio. Mentre continuavo ad ascoltare il *Decameron*, mi resi conto che Boccaccio aveva effettivamente colmato il divario tra il mondo medievale e quello moderno. Le sue storie erano ambientate nel passato, ma parlavano di esperienze ed emozioni umane universali che trascendevano il tempo e il luogo. Ha usato la sua conoscenza del passato per comprendere e dare un senso al presente e per immaginare un futuro migliore. Boccaccio non era solo uno scrittore, ma un visionario che ha visto il potenziale della letteratura come forza positiva nel mondo. La sua passione per i classici e la sua fede nel potere degli esseri umani di creare e ispirare è qualcosa che risuona con me fino ad oggi. Quindi, mentre tornavo ogni giorno al mio corso di storia europea AP, ho sentito un rinnovato senso di scopo. Non vedo più la storia come un argomento arido e noioso, ma come un arazzo di storie umane, lotte e trionfi. E sapevo che Boccaccio, con la sua passione per gli antichi e la sua fede nel valore dell'essere umano, aveva contribuito ad aprirmi gli occhi sulle meraviglie del passato e del presente. Attraverso l'infatuazione iniziale, mi sono innamorata del *Decameron*. Senza questa esperienza, non avrei potuto concepire questo progetto. Non avrei avuto la curiosità di usare la stilometria e l'elaborazione del linguaggio naturale per ricercare come le ballate del *Decameron* rivelino la psiche emotiva della brigata e aggiungano ordine alle strutture narrative del romanzo. Non avrei avuto l'opportunità di leggere (e presto cantare) ballate che non dimenticherò mai.

## **Appendice**

**Collegamento Github a tutti i file utilizzati in questo progetto:**

**<https://github.com/nkhumalo/DecameronThesis>**

## Riferimenti

- “2. He Said, She Said, We Read: An Ethical Reflection on a Confluence of Voices.” In *Ethical Dimension of the “Decameron,”* 29–38. University of Toronto Press, 2015.  
<https://doi.org/10.3138/9781442625754-005>.
- “BallateDelDecameron[19709].Pdf,” n.d.
- Beck, Eleonora. *Medievalia et Humanistica. N.S., 24: Historical Inquiries / Ed. by Paul Maurice Clogan*, 1997.
- Bent, Ian D. Review of *Review of Music from the Time of Boccaccio’s Decameron, Morrow, Beckett*, by Musica Reservata, Morrow, and Beckett. *The Musical Times* 111, no. 1527 (1970): 513–513. <https://doi.org/10.2307/956046>.
- Boccaccio, Giovanni, and Vittore Branca. *Decameron*. Einaudi, 2016.
- Brown, Howard Mayer. “Fantasia on a Theme by Boccaccio.” *Early Music* 5, no. 3 (July 1977): 325–41. <https://doi.org/10.1093/earlyj/5.3.325>.
- C. GIUNTA, *Versi a un destinatario. Saggio sulla poesia italiana del Medioevo*. Accessed October 6, 2022.  
[https://www.mulino.it/isbn/9788815084446?forcedLocale=it&fbrefresh=CAN\\_BE\\_ANY\\_THING](https://www.mulino.it/isbn/9788815084446?forcedLocale=it&fbrefresh=CAN_BE_ANY_THING).
- Cantor, Norman F. *In the Wake of the Plague: The Black Death and the World It Made*. Harper Collins, 2002.
- Cervigni, Dino S. “Fiammetta’s Song of Jealousy: Are The Young People Still at Play?” *Annali d’Italianistica* 31 (2013): 459–507.
- . “The Decameron’s Ballads and Emilia’s Happy Song.” *Annali d’Italianistica* 31 (2013): 131–71.
- Ciabattoni, Francesco. “Music in Trecento Italy and the Soundtrack of Boccaccio’s Decameron.” *MLN* 134, no. S (2019): S-138-S-151. <https://doi.org/10.1353/mln.2019.0063>.
- Clarke, K. P. “A Good Place for a Tale: Reading the ‘Decameron’ in 1358-1363.” *MLN* 127, no. 1 (2012): 65–84.
- Cooper, A., Maria Antoniak, Christopher De Sa, Marilyn Migiel, and David Mimno. “‘Tecnologica Cosa’: Modeling Storyteller Personalities in Boccaccio’s ‘Decameron’.” In *Proceedings of the 5th Joint SIGHUM Workshop on Computational Linguistics for Cultural Heritage, Social Sciences, Humanities and Literature*, 147–53. Punta Cana, Dominican Republic (online): Association for Computational Linguistics, 2021.  
<https://doi.org/10.18653/v1/2021.latechclfl-1.17>.
- Cottino-Jones, Marga. “The Mode and Structure of Tragedy in Boccaccio’s Decameron (IV, 9),” n.d., 27.
- Daelemans, Walter. “Explanation in Computational Stylogometry.” In *Computational Linguistics and Intelligent Text Processing*, edited by Alexander Gelbukh, 7817:451–62. Lecture Notes in Computer Science. Berlin, Heidelberg: Springer Berlin Heidelberg, 2013.  
[https://doi.org/10.1007/978-3-642-37256-8\\_37](https://doi.org/10.1007/978-3-642-37256-8_37).
- “Decameron Ital.Pdf,” n.d.
- “Decameron Web | Arts.” Accessed October 6, 2022.  
[https://www.brown.edu/Departments/Italian\\_Studies/dweb/arts/](https://www.brown.edu/Departments/Italian_Studies/dweb/arts/).
- “Decameron Web | Arts.” Accessed October 6, 2022.  
[https://www.brown.edu/Departments/Italian\\_Studies/dweb/arts/music/](https://www.brown.edu/Departments/Italian_Studies/dweb/arts/music/).

- “Decameron Web | Arts.” Accessed October 6, 2022.  
[https://www.brown.edu/Departments/Italian\\_Studies/dweb/arts/music/muprelu.php](https://www.brown.edu/Departments/Italian_Studies/dweb/arts/music/muprelu.php).
- Eder, Maciej. “Visualization in Stylometry: Cluster Analysis Using Networks.” *Digital Scholarship in the Humanities* 32, no. 1 (April 2017): 50–64.  
<https://doi.org/10.1093/llc/fqv061>.
- Falvo, Joseph. “Ritual and Ceremony in Boccaccio’s ‘Decameron.’” *MLN* 114, no. 1 (1999): 143–56.
- Halteren, Hans van, Harald Baayen, Fiona Tweedie, Marco Haverkort, and Anneke Neijt. “New Machine Learning Methods Demonstrate the Existence of a Human Stylome.” *Journal of Quantitative Linguistics* 12, no. 1 (April 2005): 65–77.  
<https://doi.org/10.1080/09296170500055350>.
- “How-to-Read-a-Dendrogram-Web-Ready.Pdf.” Accessed March 16, 2023.  
<https://wheatoncollege.edu/wp-content/uploads/2012/08/How-to-Read-a-Dendrogram-Web-Ready.pdf>.
- Irwin, Bonnie D. “ORAL TRADITION 10.1 - What’s in a Frame? The Medieval Textualization of Traditional Storytelling,” n.d.
- Jensen, George H., and John K. DiTiberio. “Personality and Individual Writing Processes.” *College Composition and Communication* 35, no. 3 (October 1984): 285.  
<https://doi.org/10.2307/357457>.
- Jockers, Matthew L. *Text Analysis with R for Students of Literature*. Quantitative Methods in the Humanities and Social Sciences. Cham: Springer International Publishing, 2014.  
<https://doi.org/10.1007/978-3-319-03164-4>.
- Kuhns, Richard. *Decameron and the Philosophy of Storytelling : Author as Midwife and Pimp*. Columbia University Press, 2005.
- . *Decameron and the Philosophy of Storytelling: Author as Midwife and Pimp*. New York: Columbia University Press, 2005.
- Matthews-Grieco, Sara F. “Sexualities, Textualities, Art and Music in Early Modern Italy: Playing with Boundaries. Melanie L. Marshall, Linda L. Carroll, and Katherine A. McIver, Eds. Farnham: Ashgate Publishing Limited, 2014. Xv + 242 Pp. \$104.95.” *Renaissance Quarterly* 68, no. 3 (2015): 1093–95. <https://doi.org/10.1086/683938>.
- Migiel, Marilyn. *The Ethical Dimension of The “Decameron.”* Toronto, UNITED STATES: University of Toronto Press, 2015. <http://ebookcentral.proquest.com/lib/bowdoin-ebooks/detail.action?docID=4669244>.
- Milner, Stephen J. “Boccaccio’s *Decameron* and the Semiotics of the Everyday.” In *The Cambridge Companion to Boccaccio*, edited by Guyda Armstrong, Rhiannon Daniels, and Stephen J. Milner, 1st ed., 83–100. Cambridge University Press, 2015.  
<https://doi.org/10.1017/CCO9781139013987.008>.
- Osood, Charles G. *Boccaccio on Poetry: Being the Preface and the Fourteenth and Fifteenth Books of Boccaccio’s Genealogia Deorum Gentilium in an English Version with Introductory Essay and Commentary*. Vol. 82. New York: The Liberal Arts Press, n.d.
- Sherman, Malcolm J. Review of *Review of Computation into Criticism: A Study of Jane Austen’s Novels and an Experiment in Method.*, by J. F. Burrows. *Journal of the American Statistical Association* 84, no. 405 (1989): 348–49. <https://doi.org/10.2307/2289908>.
- “Singing in the Garden: An Examination of Music in Trecento Painting and Boccaccio’s ‘Decameron’ - ProQuest.” Accessed October 6, 2022.  
<https://www.proquest.com/docview/304066764/abstract/ADC3973F44B84A3CPQ/1>.



- Vajjala, Sowmya. "4. Text Classification - Practical Natural Language Processing [Book]." Accessed December 21, 2022. <https://www.oreilly.com/library/view/practical-natural-language/9781492054047/ch04.html>.
- Vespaziani, Alberto. "'Force Disguised as Reason': Law, Jurists, and Constituent Power in Boccaccio's *Decameron*." *German Law Journal* 19, no. 4 (July 1, 2018): 879–900. <https://doi.org/10.1017/S2071832200022902>.
- Weber, Samuel. "Ambivalence, the Humanities and the Study of Literature." *Diacritics* 15, no. 2 (1985): 11–25. <https://doi.org/10.2307/464978>.